OPEN ACCESS
IOP Publishing
Modelling and Simulation in Materials Science and Engineering
Modelling Simul. Mater. Sci. Eng. **29** (2021) 085006 (38pp) https://doi.org/10.1088/1361-651X/ac2b31

Development of data-driven *spd* tight-binding models of Fe—parameterisation based on QSGW and DFT calculations including information about higher-order elastic constants

Bartosz Barzdajn[1,*], Alexander M Garrett[1], Thomas M Whiting[2] and Christopher P Race[1]

[1] Department of Materials, The University of Manchester, United Kingdom
[2] Department of Materials, Imperial College London, United Kingdom

E-mail: bartosz.barzdajn@manchester.ac.uk

Received 9 June 2021, revised 3 September 2021
Accepted for publication 22 September 2021
Published 20 October 2021

Abstract

Quantum-mechanical (QM) simulations, thanks to their predictive power, can provide significant insights into the nature and dynamics of defects such as vacancies, dislocations and grain boundaries. These considerations are essential in the context of the development of reliable, inexpensive and environmentally friendly alloys. However, despite significant progress in computer performance, QM simulations of defects are still extremely time-consuming with *ab-initio*/non-parametric methods. The two-centre Slater–Koster (SK) tight-binding (TB) models can achieve significant computational efficiency and provide an interpretable picture of the electronic structure. In some cases, this makes TB a compelling alternative to models based on abstraction of the electronic structure, such as the embedded atom model. The biggest challenge in the implementation of the SK method is the estimation of the optimal and transferable parameters that are used to construct the Hamiltonian matrix. In this paper, we will present results of the development of a data-driven framework, following the classical approach of adjusting parameters in order to recreate properties that can be measured or estimated using *ab-initio* or non-parametric methods. Distinct features include incorporation of data from QSGW (quasi-particle

*Author to whom any correspondence should be addressed.

self-consistent GW approximation) calculations, as well as consideration of higher-order elastic constants. Furthermore, we provide a description of the optimisation procedure, omitted in many publications, including the design stage. We also apply modern optimisation techniques that allow us to minimise constraints on the parameter space. In summary, this paper introduces some methodological improvements to the semi-empirical approach while addressing associated challenges and advantages.

Keywords: self-consistent tight-binding, iron, point-defects, global-optimisation

(Some figures may appear in colour only in the online journal)

1. Introduction

1.1. Motivation

We will present the development of our tight-bonding parametrisation methodology in the context of its applications. Our long-term objective is to investigate diffusion mechanisms in low-alloy reactor pressure vessel steels and to understand precipitates and nano-cluster formation. These types of defects are important in the context of life-time studies of irradiated structural materials. A good model for this class of problems has to be able to approximate well enough formation and migration energies of point defects (vacancy and self-interstitial) while being able to predict the interactions of solute atoms with these defects.

While such investigations are usually conducted using density-functional theory (DFT), they require significant computational resources, especially if we would like to consider multiple defects or dislocations and explore the whole range of possibilities. Classical empirical potentials are an excellent interpolator, but require large data-sets to optimise and have limited predictive value. Among quantum-mechanical models, the two-centre Slater–Koster (SK) tight-binding (TB) method ([1]) provides one of the simplest one-particle bases ([2]), namely the linear-combination of atomic orbitals (LCAO), with well-established rules to construct the Hamiltonian for *s*, *p* and *d* valence elements. This results in significantly increased computational efficiency and, for this reason, this classical technique is of ongoing interest within the material science community. However, the biggest associated challenge is to find transferable ('universal') parameters for constructing Hamiltonian matrix elements.

Here we revisit this problem as an automated, simple, and data-driven method for finding parameters which we believe would be an asset to the community. We will explore one of the potential approaches which is similar to, among others, the Naval Research Laboratory Tight-Binding methodology ([3]). In this approach hopping integrals are adjusted to recreate features of the electronic structure, while parameters of the repulsive pair-potential are optimised for mechanical properties.

1.2. Context

We now address the problem of model selection. There are two major reasons for presenting an alternative parametrisation for Fe, a material which has been studied in the past. First, we are interested in studying low-alloy steels in the future. Secondly, Fe, as a ferromagnetic material, serves as a good case study for development of our techniques. We are trying to achieve a balance between the predictive value of the model, speed and also the simplicity of the fitting process. In other words, we would like to find a convenient and partially automated method of parametrisation.

As a ferromagnetic transition metal, Fe requires appropriate treatment of electronic correlations. The challenge associated with the treatment of this phenomenon is best illustrated by Mehl *et al* ([4]), where a methodology that worked well for non-magnetic metals, failed in case of Fe, resulting in some elastic constants being negative. However, these correlations are not as 'troublesome' as in the case of highly correlated systems that involve e.g. f-valence electrons. Methods based on the most common implementations of the density functional theory (Khon–Sham DFT, [5]) are fairly successful with 3D transition metals. This suggests that self-consistent TB methods should be able to handle these type of materials as there is a direct link between such variants of the TB method (more information will be provided in the following section) and DFT with LDA (local density approximation) exchange–correlation functional [6].

In both methods (TB and DFT), two-electron states, responsible for the description of electronic correlations, are projected into a one-electron basis at each step of a self-consistency cycle. On the other hand, it is also well known that appropriate description of the bond-lengths requires, in the case of Fe, a generalised gradient approximation (GGA) to the exchange–correlation functional. This requirement might indicate that the TB picture is incomplete and, in some sense, the lack of physics needs to be compensated with adjustments of the model parameters. In our research we decided to use the polarisable-ion self-consistent TB model [6]. As it goes beyond a simple Stoner model ([7]), we hope to achieve a good representation of magnetism and associated effects.

Now we are left with two important choices: the basis and method of estimating transferable parameters. Examples of successful models of Fe include: a non-orthogonal *spd* basis by Bacalis *et al* [8] (no self-consistency), an orthogonal d-band by Liu *et al* [9], a self-consistent non-orthogonal *spd* model by Paxton and Finnis [6], another similar model with s–d basis by Paxton and Elsässer [10], optimised orthogonal d-band by Madsen *et al* [11], and self-consistent DFT-based orthogonal d-band model by Hatcher *et al* [12]. DFT-based parametrisation of the SK-TB has a great advantage as it addresses the main issue of finding suitable parameters (see also Horsfield and Kenny [13] for more information on DFT-TB). However, DFT itself does not necessarily provide an optimal approximation and its quality may worsen when an LCAO basis is used. On the other hand, this can be compensated by refitting. Hence, we decided to focus our efforts on the semi-empirical scheme, as it can be used for refinement or as a stand-alone technique. An additional practical advantage is that the developed framework can then be used with any atomistic software package that implements the TB scheme. To minimise constraints we will be using a full *spd* basis, which will allow us also to recreate most of the features of the band structure.

1.3. Methodology

Finding TB parameters when using a full *spd* non-orthogonal basis is a significant challenge. One of the key features of this paper is a presentation of the methodology used in the formulation of the optimisation problem. This includes the reasoning behind selecting fitness measures and objectives. For example, as an objective, we have used band structures estimated using three different methods: DFT with LDA, DFT with GGA and the quasi-particle self-consistent GW method—QSGW ([14]). To be more specific, we have used band structures generated from the density updated in the QSGW self-consistency cycle [15]. We will demonstrate that it is possible to obtain a very good agreement with the QSGW results. This might be of particular interest for the community, as QSGW provides high-quality truly *ab-initio* results. We believe, that this approach may be used to improve existing parametrisations with results from more complete (in the physical sense) models. Here, we are referring to the more accurate *ab-initio*

treatment of electronic correlations. While the DFT implemented with the GGA works very well in case of Fe, the same approach may be adopted for materials for which it is not the case.

Regarding our approach to selection of TB parameters, it would not be possible to obtain such a good agreement with reference data without using a somewhat novel approach, i.e. usage of multiple measures, each approaching a different optimum, together with an efficient global optimisation algorithm. In this way we reduced issues associated with ill-posed optimisation problems.

We also show that by aiming to recreate third-order elastic constants we can sufficiently sample pair-wise interactions without applying large strains (large strains can move some atoms outside their interaction range). This is another way we reduce the degree to which the problem is ill-posed as we exclude test-cases with a different number of considered neighbours. Also, we show that some deformation patterns can result in negative curvature in the energy–strain relationship even when other deformation patterns give reasonable elastic constants. By applying additional test-cases, necessary for the calculation of higher order elastic constants, we increase the robustness of the optimisation.

Additionally, we demonstrate that using energy–strain curves estimated using the DFT method we can recreate experimental elastic constants. This approach has the advantage of being a much more computationally efficient way to quantify the goodness of fit. Furthermore, second derivatives are extremely sensitive to changes in model parameters and tend to make optimisation much more difficult. Finally, we were able to satisfy the objectives and obtain a reasonable transferability, while applying as little constraint as possible on parameters. However, the predictive value of our models, here represented by formation energies of point defects—our main quantity of interest, may benefit from further improvement.

2. Description of the model

The TB approximation to the many-body problem, which we consider in this paper, is a very well established technique. However, the name serves also as an umbrella term for many somewhat distinct methods. There are also many ways various effects can be included in models from the same category. For a general review we recommend publications by Paxton [16], Sutton and Balluffi [17] or Finnis [18].

The model we use is the self-consistent polarisable TB model, as presented by Paxton and Finnis in [6] (see also Finnis *et al* [19] and Sutton *et al* [20]), as implemented in the Questaal package [21, 22]. The theory behind the TB approximation we will address only briefly, in order to emphasise the key physical effects that are included and how this is done. The binding energy E_{B} (total energy minus energy of isolated atoms) of a system within this approximation, can be written as a sum [17],

$$E_{\mathrm{B}} = E_{\mathrm{bond}} + E_{\mathrm{prom}} + E_{\mathrm{pair}}, \tag{1}$$

where E_{bond} is the covalent bond energy, E_{pair} represents Coulomb interactions between nuclei and E_{prom} represents changes in the energy due to differences in electronic configuration of a solid compared to isolated atoms. As argued by Sutton *et al*, E_{pair} may also encapsulate changes in electrostatic energy due to exchange–correlation interactions [17, 20]. The approach adopted by e.g. Paxton *et al* differs in this context. The TB Hamiltonian associated with the sum $E_{\mathrm{bond}} + E_{\mathrm{prom}}$ is decomposed into two components

$$\hat{H} = \hat{H}_1 + \hat{H}_2, \tag{2}$$

where $\hat{H}_1$ represents non-interacting electrons in an effective potential and $\hat{H}_2$ '… describes electron–electron interactions and is constructed so as to represent second order terms in the expansion of the Hohenberg–Kohn density functional about a reference density …' [6]. In this approach less is required of the pair-potential in reproducing the total energy.

The model for electrons can be summarised more explicitly, yet in a fairly compact way, using a projection to a Hubbard-like Hamiltonian [23]. Using the formalism of the second quantisation and LCAO basis (e.g. [24, 25])

$$\hat{H} = \sum_{\sigma}\sum_{\alpha\beta} h_{\alpha\beta\sigma}\hat{c}^{\dagger}_{\alpha\sigma}\hat{c}_{\beta\sigma} + \sum_{\sigma\sigma'}\sum_{\alpha\beta\gamma\delta} \mathcal{U}_{\alpha\beta\gamma\delta\sigma\sigma'}\hat{c}^{\dagger}_{\alpha\sigma}\hat{c}_{\beta\sigma'}\hat{c}^{\dagger}_{\gamma\sigma'}\hat{c}_{\delta\sigma}, \tag{3}$$

where $\hat{c}^{\dagger}$ and $\hat{c}$ are creation and annihilation operators respectively. In the above it is assumed that all normalisation factors are included in the matrices. The Greek letters represent a combined index of site positions $\vec{R}$ as well as quantum numbers l and m, while σ indexes the spin. The first term in equation (3), represents elements of $\hat{H}_1$ i.e. the classical non-self-consistent TB approximation. The second term, corresponds to H_2 and defines pair-wise Coulomb interactions. In the spirit of the Hubbard model, these are limited to intra-atomic interactions (see [24]) and the matrix $\hat{\mathcal{U}}$ is non-zero only for diagonal and off-diagonal on-site elements. It consists of pair-wise terms, such as Coulomb integrals (for combinations of a type $\alpha\alpha\beta\beta$), and exchange ($\alpha\beta\beta\alpha$ type) only for atomic orbitals originating from the same atom. Inclusion of spin changes the span of $\hat{H}$ to include the basis of Pauli matrices.

In the given implementation, electron–electron contributions are included in a self-consistent manner, by contributing to all (diagonal and off-diagonal) on-site elements. Self-consistent interactions include Madelung and Hubbard potentials that yield appropriate shifts to the non-interacting Hamiltonian. The exchange is included via the Stoner parameter. The cycle in the self-consistency loop continues until input and output charge distributions are self-consistent. The TB model can be regarded as a single-particle picture because within each iteration the model is projected into a one-particle basis. The basis set is defined in keeping with the SK algorithm as are the rules for the creation of the initial Hamiltonian matrix [1]. The core idea behind the SK approximation is to represent Hamiltonian matrix elements (H_{1ij}) in terms of two-centre integrals. These integrals are given as

$$H_{1ij} = E_{\alpha\beta} = \int_V \phi_\alpha\left(\vec{r} - \vec{R}_I\right)\hat{H}_1\phi_\beta\left(\vec{r} - \vec{R}_J\right)\mathrm{d}\vec{r}, \tag{4}$$

where $\phi_{\alpha(\beta)}$ are atomic orbitals, $\hat{H}_1$ is the Hamiltonian $\hat{H}_1 = \hat{T} + \hat{V}_{\mathrm{eff}}$, consisting of the operator of kinetic energy of electrons $\hat{T}$ and effective potential $\hat{V_{\mathrm{eff}}}$—approximated as a sum of spherically symmetric contributions centred at each atom of the unit cell. These integrals represent contributions to the bond from distinct pairs of atomic orbitals. In the Slater and Koster algorithm three-centre contributions are ignored. By setting the coordinate system for each pair in such a way that the z axis originates on one atom and connects with the other, it is possible to rewrite Hamiltonian matrix elements in terms of direction cosines and (fundamental) bond integrals, also referred to as hopping integrals. For each atomic species, we need to find values of these integrals and their dependence on the bond length. In case of the self-consistent TB, the parameters define initial values of hopping integrals that are later updated during the cycle. The procedure continues until forces are self-consistent. Effects included in this procedure have been discussed in the previous paragraph. For more details we refer the reader to already cited works by Finnis *et al* ([19]) and Paxton and Finnis [6]. As emphasised earlier, this method provides a significant computational efficiency while including a representation of the electronic correlations.

2.1. The issue of parameter space dimensionality

In the case of the *spd* basis, there are 9×9 possible orbital pairs. In the two-centre approximation, due to the symmetry, this number reduces to 29 that need to be specified, which further can be represented using only 10 fundamental bond integrals (table 1 in Slater and Koster [1]). These integrals are designated as: $ss\sigma$, $sp\sigma$, $pp\sigma$, $pp\pi$, $sd\sigma$, $pd\sigma$, $pd\pi$, $dd\sigma$, $dd\pi$ and $dd\delta$. The first two Latin letters in this naming convention represent the type of orbitals that contribute to the bond, while the Greek letters correspond to the bond type. More information can be found in the original SK paper [1]. Applying a non-orthogonal basis, with a single parameter controlling the decay, doubles the degrees of freedom of the model. We also need to find diagonal on-site elements, one for each orbital type. The Hubbard-like U can be assumed constant (1Ry is a common value), although the Stoner parameter needs to be optimised. For parameters associated with the electronic structure, we apply single-parameter exponential decay. In summary, with 10 orbitals in an *spd* basis, this gives us 40 parameters describing hopping and overlap as there are two parameters (magnitude and decay) that define a single quantity (overlap or hopping). Including the Stoner parameter and on-site energies results in 44 degrees of freedom (DOF) in total (with respect to the electronic structure). Additionally, we need to find parameters describing atomic pair-wise repulsion. This adds at least two extra DOF. Parameters can be found by comparing the results of TB calculations with experiments or *ab-initio* calculations. We are aiming to optimise a self-consistent model, based on the variational principle, on a non-orthogonal basis that trades off flexibility (completeness) for simplicity (LCAO rather than Slater determinant). This means that whatever relationship we aim to reproduce or whatever performance measure we apply, the optimisation problem is likely to be ill-posed i.e. it might be impossible to select the right constraints for parameters before optimisation is complete. In other words, depending on the objective and constraints, either there will be no optimal solution (over-constrained problem) or the solution will not be unique (under-constrained). Hence, one needs to carefully choose performance measures, objectives and stopping criteria.

At this point, we need to clarify that in our approach we differentiate between a *measure of performance* and an *objective*. The first quantity (later we will refer to it also as a loss function) simply reflects the quality of a candidate solution, and several of them will be tracked during the optimisation. The objective will be the one performance measure selected to guide the optimisation process.

3. Defining performance measures for band structure parameters

Band structures are inexpensive to calculate while providing a significant amount of information about the electronic structure. This is an important characteristic that will allow us to use less restrictive constraints on TB parameter space and explore a wider range of candidate solutions. The selection of candidate solutions will be based on different performance measures. The reasoning is as follows. Each distinct measure will define a different optimisation landscape and it will guide the algorithm to the optimal solution through a different path. Even if we consider a family of measures that would point to the same solution in the case of a perfect fit, due to intrinsic properties of the TB approximation and sub-optimal boundary conditions, achieving this will be either impossible or very unlikely. Hence, the usage of distinct measures will result in a whole set of candidates leading to a greater confidence in the final result.

We define the loss/cost function in a way analogous to the Minkowski distance i.e.

$$f_{\mathrm{B}}\left(\vec{\alpha}_{\mathrm{B}}\right)=\sum_{s\sigma}\left(A\left(s,\sigma\right)\sum_{nk}\left|\epsilon_{nk}^{(\mathrm{TB})}\left(\vec{\alpha}_{\mathrm{B}},s\right)-\epsilon_{nk}^{(\mathrm{ref.})}\left(s\right)\right|^{p}\right)^{1/p'}, \tag{5}$$

where $\vec{\alpha}_{\mathrm{B}}$ is a tuple representing all parameters associated with the electronic structure, while $\boldsymbol{\epsilon}^{(\mathrm{TB})}\left(\vec{\alpha}_{\mathrm{B}},s\right)$ and $\boldsymbol{\epsilon}^{(\mathrm{ref.})}\left(s\right)$ are matrices of TB and reference band structures (arranged in a matrix of eigenvalues) respectively. The parameter s denotes both crystallographic structure and volumetric strain γ. Possible values can be represented by a Cartesian product

$$s\in\left\{\mathrm{bcc},\mathrm{fcc},\mathrm{hcp}\right\}\times\left\{\gamma_1,\gamma_2,\gamma_3\right\}. \tag{6}$$

The sum in (5) is carried over all elements of the product from equation (6). The dominant interactions will be between first neighbours. With a single parameter decay, a set of three strains is the second smallest that resolves ambiguity arising from the dependence on the bond length.

As for other parameters, σ indicates spin while A is a normalisation function that assigns higher weight to the majority spin. Additionally, the latter compensates for differences in number of eigenvalues in hcp bands. The sum is carried over all eigenvalues (index n) and samples of the Brillouin zone (index k). These details are relevant as they define the shape of the loss function and an optimal solution as a result. This is due to fact that it is unlikely we will be able to achieve a 'perfect' fit.

We distinguish between exponents p and p' so we can control the shape of a sphere with equally scored solutions and the shape of the objective function separately. Other measures, such as the Jaccard similarity coefficient, Canberra distance or total variational distance, were also tested. However, the measure (5) proved to be the most successful. It is flexible in terms of shaping of the loss function landscape and can handle missing data or NaN values well. This is a critical property as some model parameters lead to an ill-posed eigenvalue problem. The summation always yields finite numbers, unless band calculations fail.

Here we would like to address the consequences of selecting different p and p' values. Researchers experienced in the development of empirical and semi-empirical models might safely ignore the reminder of this section. Exponent p' is very straightforward to interpret as it simply corresponds to raising the 'base' loss function to a specific power. As a result, to some extent, we can control the balance between exploration and exploitation when evolutionary algorithms are in use. In other words, higher values of the exponent will promote faster convergence to minima (local and global) at the cost of exploration of the objective landscape. In our experiments we usually used values $p'=p$ in initial stages and $p'=1$ for more aggressive refinement.

Figure 1 illustrates how changes in the value of p result in different 'preferences' of the loss function. In this example, we consider a simple case of comparing values of two functions, a candidate and a reference (target), denoted by $y^{\mathrm{can}}(x)$ and $y^{\mathrm{ref}}(x)$ respectively. Assume we evaluate these functions at two points of interest, x_1 and x_2 and measure the distance between the points/vectors $(y^{\mathrm{can}}(x_1),y^{\mathrm{can}}(x_2))$ and $(y^{\mathrm{ref}}(x_1),y^{\mathrm{ref}}(x_2))$ according to different values of the exponent p. As the form of the candidate function varies the deviations from the values of the reference function at the two points x_1 and x_2 will change. We denote these deviations $\Delta y_i=y_i^{\mathrm{ref}}-y_i^{\mathrm{can}}=y^{\mathrm{ref}}(x_i)-y^{\mathrm{can}}(x_i)$ for $i=1,2$. This is equivalent to comparison of two vectors in two dimensional space. The first row of plots in figure 1 shows contours of equidistance between these two points for different values of p—essentially a contour along which $y^{\mathrm{can}}(x)$ is an equally good fit to $y^{\mathrm{ref}}(x)$ given its values at x_1 and x_2. For $p=1$ the trade-off between a good fit at one point and that at the other is 'even-handed', whereas for increasingly high values of p small deviations from a good fit at x_1 must be offset by larger improvements of

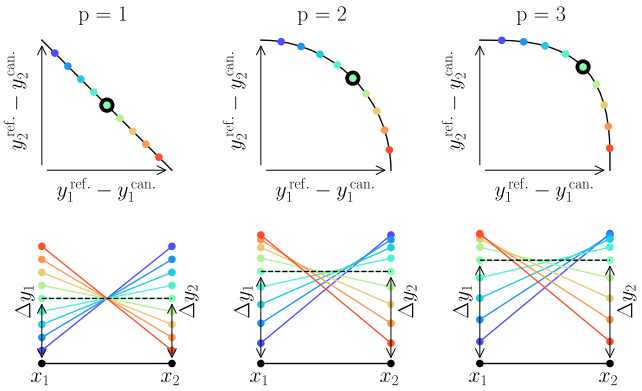

Figure 1. The upper plots illustrate the effect of different values of the exponent p of the Minkowski distance on contours of equal goodness-of-fit of a candidate function $y^{\text{can}}(x)$ to a reference $y^{\text{ref}}(x)$ when evaluated at two points x_1 and x_2. In the lower plots, the trade-off between deviations in fit at the two points is shown for points equally distributed along the contours in the upper plots. Colours of the points and lines indicate the correspondence between upper and lower plots with the horizontal dashed line corresponding to the circled point of equal deviation from the reference at the points x_1 and x_2.

the fit at x_2 and vice versa. The second row of figure 1 illustrates this idea in another way, representing the balance between changes in Δy_1 and Δy_2 (i.e. in the goodness of fit at the two points) for different values of p for points evenly distributed along a loss contour. In other words, starting with $p = 1$, as we increase the value of p we move from the so-called Manhattan distance towards Tchebychev distance. For values, $p = 3$ and higher an equivalent metric could be regarded as a soft-maximum metric, although this is highly subjective.

As mentioned before, these considerations are important in the context of fitting ill-posed problems. The best achievable solution will depend on the accepted objective function as solutions based on different measures will approach the optimum differently. While it is impossible to say which values are preferable, this way we can achieve diversification of candidate solutions and minimise the risk of omitting global minima. Assuming such a minimum exists.

4. Defining a performance measure for pair potentials

To obtain a good estimate of an optimal pair-potential it is necessary to evaluate changes in total energy under different lattice deformations. Usually, this is done by evaluation of elastic constants. However, calculation of derivatives at a specific point introduces a certain level of ambiguity. Additionally, the values of elastic constants are extremely sensitive to even the smallest changes. For this reason our evaluation of mechanical properties will use a similar measure to 5

$$f_{\text{pp}}\left(\vec{\alpha}_{\text{pp}}\right) = \left(\sum_{\boldsymbol{\eta}} \left| \frac{f_{\text{tb}}\left(\vec{\alpha}_{\text{pp}}, \boldsymbol{\eta}\right) - f_{\text{ref.}}\left(\boldsymbol{\eta}\right)}{f_{\text{ref.}}\left(\boldsymbol{\eta}\right)} \right|^{p} \right)^{1/p'}, \tag{7}$$

where $\vec{\alpha}_{\text{pp}}$ is a tuple of pair-potential parameters, $\boldsymbol{\eta}$ is the strain tensor and f represents the elastic enthalpy for the tight-binding model (tb) and the reference (ref.). The measure (loss function) for the pair-potentials has the distinct feature of being normalised to emphasise the

Table 1. Components of the Voigt strain used to construct deformation tensors that defines lattice deformations.

Pattern/component	η_1	η_2	η_3	η_4	η_5	η_6
A1	γ	0	0	0	0	0
A2	γ	γ	0	0	0	0
A3	γ	$-\gamma$	0	0	0	0
A4	0	0	0	0	0	2γ
A5	γ	0	0	2γ	0	0
A6	γ	0	0	0	2γ	0
A7	γ	γ	γ	0	0	0
A8	0	0	0	2γ	2γ	2γ

importance of results near equilibrium crystal structure. At this point, we would like to reiterate the distinction between the objective function and the loss function. The objective function will be the target and a measure of 'success', while the loss function will be one that is used to drive the optimisation. The proposed approach relies on *ab-initio* (non-parametric) calculations, instead of on experimental values, although the latter are values we are aiming to recreate. However, in the case of ill-posed and non-linear problems sometimes by aiming towards a less optimal solution we can get 'closer' to the objective.

We note that in principle, it would be possible to recreate energy–strain curves using higher-order elastic constants. However, available measurements for temperatures close to absolute zero have an unspecified uncertainty and repeatability seems to be quite limited. This argument will be clarified in section 9.

5. Summary of the selected performance measures

Deformations of a perfect crystal or one that contains defects should be sufficient to find all parameters of the model. However, these calculations are expensive because of the high dimensionality of the problem. On the other hand, we can evaluate the majority of parameters that are related to the electronic structure using fairly fast band-structure calculations. Furthermore, for evaluation of pair-potentials, we introduced a stable measure (less ambiguous and not overly sensitive) that includes information about higher-order elastic constants. Higher-order contributions will mainly depend on strains included in the loss function. Finally, definitions (5) and (7) allow control over the family of acceptable solutions as well as the shape of the objective function (by controlling p and p' parameters). Furthermore, a combination of measures should provide an optimal amount of information with respect to computational time. Another way to look at these parameters is to consider p as controlling the penalty for a maximum difference between eigenvalues across the Brillouin zone, while p' controls the penalty for poorly chosen model parameters.

6. Selecting objectives

In this section, the objective is to define a set of model predictions that will be combined using the above measures to give a numerical representation of the fitness/loss. To provide some overall context, consider a set of deformations necessary to represent higher-order elastic constants of cubic materials. These deformation patterns are given in table 1. Patterns A1–A6 are based on ones given in [26]. Additional patterns A7 and A8 are necessary to resolve c_{123} and c_{456}

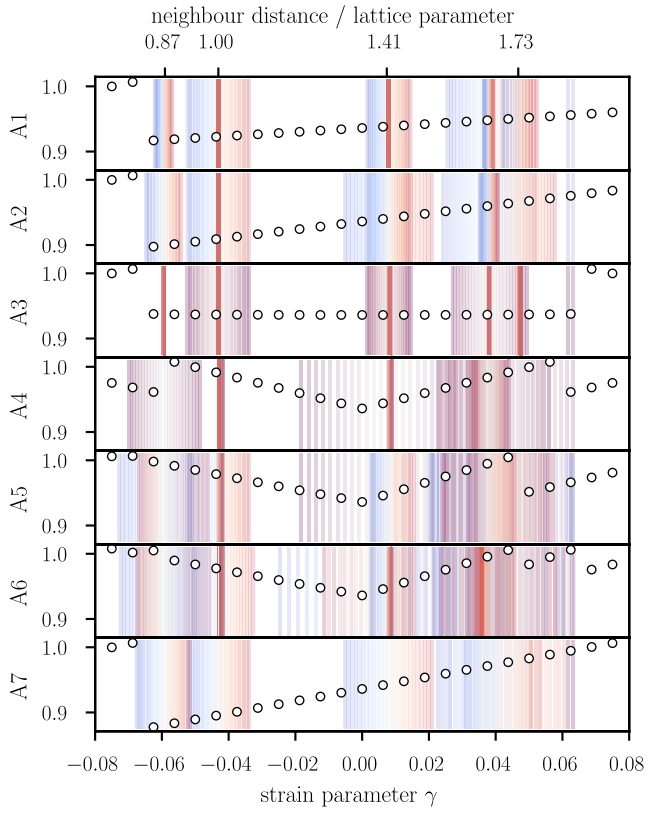

Figure 2. Evolution of distances, measured on the upper horizontal axis, between pairs of atoms (vertical lines) under deformations of a bcc lattice presented in table 1 (excluding deformation A8). Blue colour indicates negative values of γ while red—positive. Scatter plots represent normalised maximum distance included within cut-off radius $2a_{\mathrm{bcc}}$, where a_{bcc} is the equilibrium bcc lattice parameter (2.866 (Å)). The lower x-axis is associated with the scatter plot, while the upper x-axis with the evolution of distances (indicated by vertical lines).

elastic constants. In explanation of our selection of objectives we begin by arguing that deformations necessary to calculate higher order elastic constants are sufficient to sample the space of local atomic environments appropriately. That is, by aiming to recreate elastic constants (up an arbitrary order and by including various crystallographic structures) it is possible to resolve all model parameters. Though this is just an assumption, we can gain some confidence in this statement by looking at the influence of various deformations on the sampling of the pair-wise interactions (figure 2), to be discussed below.

However, while deformations in table 1 are necessary to estimate elastic constants, some of them are not necessarily the best choice for evaluation of how well a model can represent band structures. This assertion is based on our experience and we argue that this is a result of emphasis on further neighbours. For example, taking deformations A1 and A3 will provide us with test cases that are different only in minor features of a band structure when model parameters are optimal. Furthermore, it is likely that we will not be able to recreate them due to model limitations. These considerations are important as evaluation of bands is an inexpensive source of high quality information about the goodness of fit. Hence, we would like to put more

Table 2. Population size of neighbour-shells in different structures.

Neighbour/structure	1st	2nd	3rd
Diamond	4	12	12
Simple cubic	6	12	8
bcc	8	6	12
hcp	12	6	2
fcc	12	6	24

emphasis on these calculations in the fitting process. Therefore we shift our focus to the pattern A7. Deformations A1–A6 will again become relevant in the context of elastic properties.

We would like to implicitly sample the relation between distance and magnitude, for hoppings and overlaps alike, across all relevant distances between pairs of orbitals. As we can see from figure 2, deformations vary significantly in the way that they affect the distribution of distances between pairs. On the other hand, the starting point for the model is non-self-consistent TB, a method that can be at best a variational approximation to the ground state, where effective potential/density is implicitly defined by hopping parameters. Therefore, the primary objective should be a recreation of band structures at equilibrium lattice parameters for realistic structures. Hence, we select test cases that maximise sensitivity to changes in parameters while not deviating too much from the ground state. By emphasising the variety of structures over a variety of deformations we are also minimising the issue of not being able to recreate changes in the reference as higher strains may result in changing the number of neighbours within the cut-off radius.

Furthermore, we found that even in the case of a simple compression or tension it is impossible to force TB to recreate a certain class of changes in band structures that DFT or QSGW predicts. This statement mainly refers to shifts of particular band-lines. While band crossings and degeneracies were maintained, we failed to find a parametrisation that would allow us to reflect the same relative shifts as in the references. We believe it is partially related to the dependence of hopping integrals on distance. This dependence is defined by a parameterised function (discussed in section 8) that is fairly constrained. However, more flexible models (involving more parameters) proved to be extremely difficult to optimise as the dimensionality of the problem is already quite significant.

The selection of the crystallographic structures is based on the information provided in table 2. The most natural candidates are close-packed structures: hcp and fcc. These are known to work well with the atomic-sphere approximation (ASA) making them an ‘easy target’ for TB as TB and ASA are very closely related [2, 27]. Needless to say the bcc structure needs also to be included in the reference data as we are optimising the model of ferromagnetic Fe. Due to the similarity between fcc and hcp, namely the same number and distances of 1st and 2nd neighbour-shells, there is a danger of bias in the reference data-set (over-fitting). Other structures included in table 2 show potential improvements in the reference.

Regardless of the selected structures we also need to include strained structures as three structures will be insufficient to estimate optimal dependence of pair-wise interactions on distance. We decided to focus on the easiest to recreate dependence on isotropic compression and tension. As illustrated in figure 2 (A7), this should provide a sufficiently rich sampling. However, we accepted this simple reference set only because decay is defined by a single parameter. A more flexible model might require a more complex objective/loss function.

The final consideration in this scope is the magnitude of strain. The plot of the maximum pair-wise distance within cut-off radius under volumetric strain (A7) demonstrates that under

high strains we change the number of orbitals included in the construction of the Hamiltonian. This is indicated by the sudden increase of the maximum distance under compression. Hence, effectively we change the nature of the model leaving us with no other choice other than to limit the strain magnitude. Otherwise, we push the optimal solution further away from what can be achieved within the TB approximation. This is one of the reasons we limit ourselves to the following volumetric strains: -0.06, 0.00, 0.06. This gives us 3×3 structures included in the loss function (5). There is also another reason for this approach that will be addressed in section 7.

The next piece of the puzzle is the recreation of the mechanical properties. Here we address why we decided to aim for deformations that are representative for higher elastic constants. In our experiments, we noticed that any subset of three patterns from table 1 can be used to calculate second-order elastic constants (SOEC). However, that does not guarantee that the bcc structure will be stabilised. From figure 2 we can see that mode A3, that explicitly gives us the c' shearing constant, emphasises other than nearest-neighbour (NN) distances. It is possible to obtain a very good estimate of elastic constants, as a result of cancelling errors, by over-estimation of interactions at NN. In many cases, this could mean a negative curvature of the energy–strain relation when using the A3 deformation pattern.

7. Optimisation strategy and its implementation

While the objective is to predict total energies associated with defects, we argue that these calculations are too expensive to be included in the reference data-set. Therefore, we focus on the measures presented above, associated with two features of the model predictions: electronic structure and elastic properties. In the latter case, it might be possible (although unlikely) that given a sufficiently large and optimised reference set, all ambiguities with respect to model parameters could be resolved. However, calculations of elastic properties can also be very expensive given the nature of the problem (non-linearity and with high-dimensionality). Furthermore, it might be impossible to assemble all results within a single objective without arbitrarily selected weighting. For example, different elastic constants can vary by orders of magnitude and also cannot collectively be compared with formation energies. Here, multi-objective optimisation is out of the question because it is simply too expensive. Therefore, we make a case for the classical approach where optimisation is based on separate, also inexpensive, band calculations and estimation of elastic constants (energy–strain curves in our case).

In other words, we focus first on the optimisation of parameters associated with hopping integrals by targeting various band structures. These calculations do not require the pair-potential to be defined. Having candidates defining one part of the total energy we move to optimisation of pair-potential parameters. This time we are targeting total energies under specific deformations that allow us also to calculate elastic constants. A flowchart that summarises and illustrates the most important steps can be found in figure 3.

The drawback of this approach is that a good fit to the band structures does not necessarily indicate that it will be possible to find a reasonable approximation to the optimal pair-potential (leading to poorly recreated elastic constants). On the other hand, we would like to avoid situations in which good candidates for hopping integrals are rejected in an optimisation step because the algorithm is exploring the wrong subspace of pair-potential parameters at the time. Nonetheless, it is necessary to combine knowledge about both sub-sets of parameters at some point of the optimisation.

We have solved this problem by using a brute force approach. The idea is to generate a large sample of candidate band-structure parameters. For that reason, we optimised parameters in

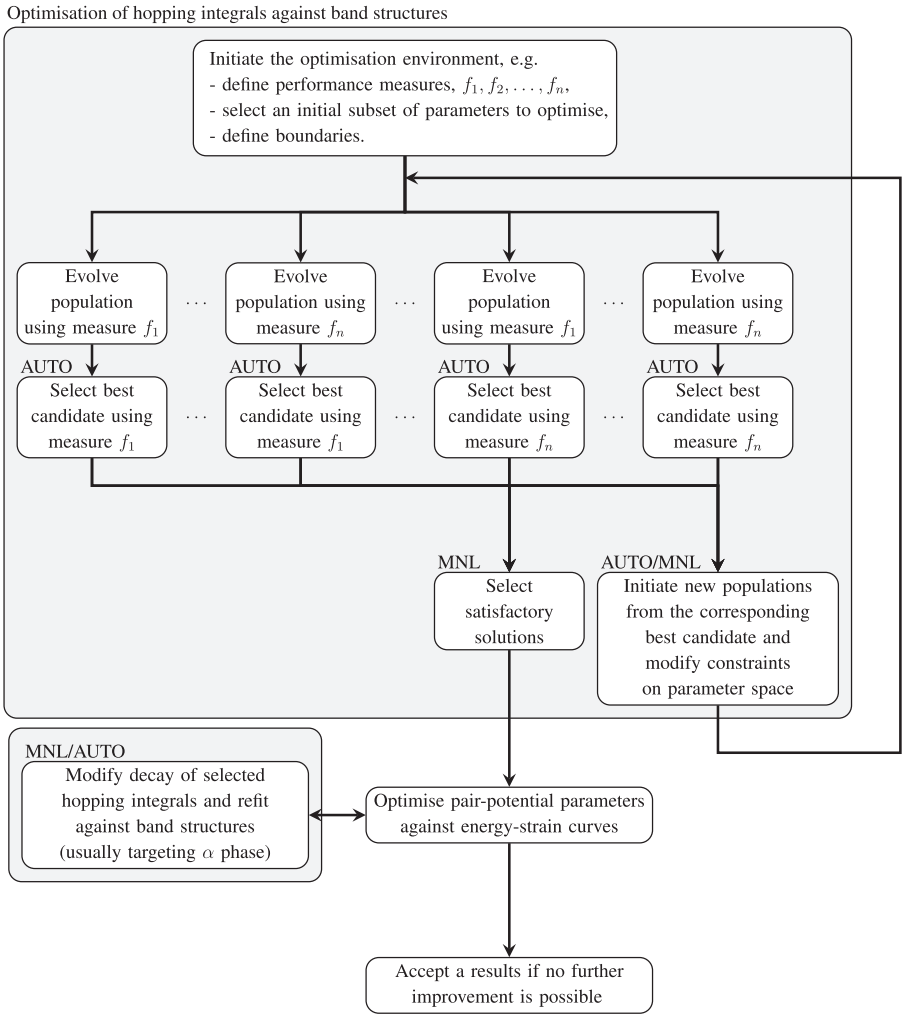

Figure 3. Flowchart of the calibration method—single reference set. There are two main stages of the process. In the first stage, we optimise hopping parameters and generate a set of candidate models based on different performance measures. Each iteration can produce parametrisations that reasonably well approximate band structures. Furthermore, with each iteration, we change the boundary conditions by making them less restrictive. In the next stage, we find optimal pair-potential for each reasonable approximation of the electronic structure. This time we are targeting various energy–strain calculations. The three best candidates were subject also to fine-tuning. In this process, we modified manually selected parameters associated with hoping integrals and refitted them. For example, we would modify decay parameters associated with *d* orbitals and attempted to readjust relevant magnitudes. Subsequently, we would refit pair-potential parameters. We focussed on maintaining a good fit to band structures while we also attempt to improve mechanical properties using the hit-and-run tactic. Unfortunately, we filed to automate this step (fine-tuning) as we have not found a convincing relationship between parameters link with the electronic structure and good mechanical properties. The grey background corresponds to stages that involved the refitting of parameters associated with hopping integrals. Abbreviations AUTO and MNL refer to automatic and manual selection processes respectively. They are given in the order of tasks described in a node.

several stages using different p-norms and references (each case was optimised separately). Parameters were fitted against DFT band-structures with LDA and PBE functionals as well as the QSGW method. Each reference was evaluated using $p \in \{1, 2, 3\}$. Subsequently, from every population (9 in total), three samples were selected based on the performance with respect to each measure. This gave us 27 candidates from a single 'epoch' (simultaneous optimisation of all populations with distinct objectives and loss measures). Starting from the previous optimum this procedure was repeated several times, giving more than 100 candidate parameterisations, in the initial stage. In the process of optimisation of band structures we generate candidates and we discriminate them on the basis of mechanical properties. Best performing candidates (parameters associated with the band structure) were used as a new starting point and the whole procedure was repeated and results subsequently refined, as will be described in the following paragraph.

In our experiments, we tested several optimisation algorithms focussing on derivative-free methods. We found that the covariance matrix adaptation evolution strategy (CMAES, [28]) was the most efficient and robust. Particle swarm optimisation (PSO) also performed well. Surprisingly, the Gaussian process optimisation with radial kernels failed to find any reasonable solution. We decided to focus on the CMAES using an implementation by Hansen *et al* [29].

Broadly speaking, in the CMAES a population of random vectors (samples) in the space of model parameters is generated from a multivariate Gaussian distribution defined by the initial mean and covariance. Subsequently, each sample is evaluated and ranked according to the loss function. On this basis, the mean and covariance matrix of the distribution is updated and the cycle is repeated. The rule-of-thumb was to set the size of the population to 10–15 times the number of degrees of freedom.

The optimisation would not be successful without setting some initial constraints on the parameters. This was implemented by forming three groups of parameters: 1: $\{dd\sigma, dd\pi, dd\delta\}$ (d-block), 2: $\{sd\sigma, pd\sigma, pd\pi\}$ (sp–d block), 3: $\{ss\sigma, sp\sigma, pp\sigma, pp\pi\}$ (sp block). The partition was inspired by the approach by Dufrense *et al* presented in [30]. Each group initially shares the decay parameters and overlaps. We placed no constraints on ratios between hopping integrals.

Following convergence of the loss function, we recorded the results as described earlier. In the next optimisation step, we modified the formulation of the problem, allowing, for example, each decay parameter to be optimised separately. The constraints were imposed and removed in each cycle after a careful examination of the features of the band structures. The idea was to build a sample of candidates that successfully recreated band structures.

At this stage, we found only basic relationships between parameters. An example can be found in figure 4. From this figure, it is clear that there is a strong correlation between the magnitude and decay among optimised parameters. Furthermore, there is a significant level of ambiguity between these quantities. This indicates that the value of the hopping integral at nearest-neighbours is a dominant factor here.

The reasoning is as follows. We assume exponential decay of hopping integrals (more details about the model parametrisation will be presented in the following section 8). As a result, we can achieve the same value at a given point by controlling the scale and decay magnitude. However, such manipulations result in significant changes elsewhere, e.g. at further distances. Given that we have multiple band structures in the reference set, the dominant role of nearest-neighbour interactions is the most likely explanation of this observation. Likewise, overlap parameters are ambiguous.

Our approach allows us to explore a wide variety of potential candidates in subsequent tests of other properties (e.g. elastic const.). However, there is no guarantee that it will be possible to find a unique solution using computationally inexpensive tests. After several cycles of optimisations with different measures and reference band structures, we attempted to fit

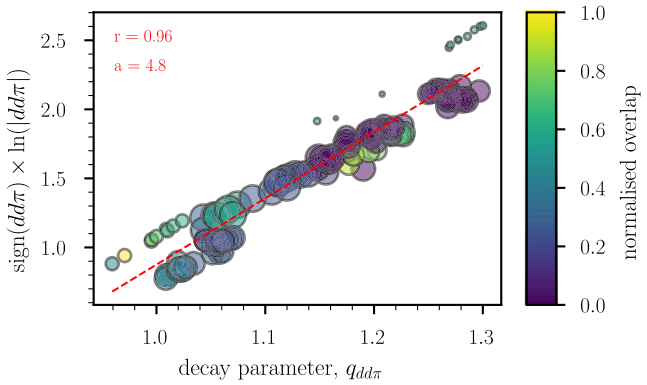

Figure 4. Example of a relationship between the loss function value, magnitude, decay and overlap. Given the exponential decay, the majority of other hopping integrals revealed a similar relationship. Size of the data-points is proportional to the value of the loss function (bigger marker indicates a better performance). The red line represents a linear fit with a value of the slope given as the variable a and correlation coefficient as r. Colours represents a scaled overlap that in case of the d-block could be considered as relatively small.

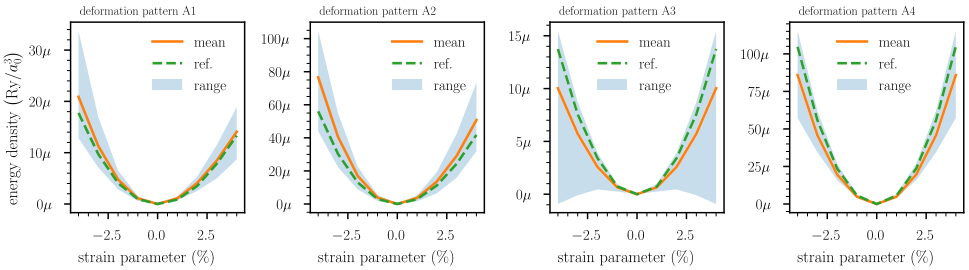

Figure 5. The mean and range of energy–strain relationships for selected deformation patterns from the reference-data-set, as estimated by candidate TB models. Note the negative values for the deformation pattern A3.

energy–strain curves and evaluate elastic constants. To speed up the process, we precalculated energies without pair potentials for each deformation pattern.

In the optimisation of pair-potentials, we used the $p = 3$ norm and recently developed by Endres *et al* [31] SHGO algorithm (with Sobol sampling) implemented in the SciPy stack [32]. We emphasise that we tested a variety of other algorithms and all failed to provide any acceptable results. The main issue was to find a solutions that would approximate the total energy reasonably well in all test cases. As indicated by summary statistics associated with selected patterns in figure 5, often we would get solutions that fit well to all patterns except A3. A very similar problem was encountered by Mehl and Papaconstantopoulos [4]. Though this also occurred with SHGO, nonetheless, this algorithm was more robust and allowed us to find a satisfactory solution with average relative differences in total energy below 20% for all tested deformation patterns.

The main danger here was that it is possible to estimate elastic constants ignoring the pattern A3 and report satisfactory predictions for an unstable crystal (negative curvature under A3 deformation). The distinct feature of this deformation pattern is weak dependence on the 1st

neighbour shell (figure 2). We argue that by including deformations that allow calculation of higher-order elastic constants we introduced a more robust framework. The drawback is that we need to fit energy–strain curves (due to high sensitivity of derivatives to model parameters). Hence, we cannot rely on experimental data.

At this point, we arrived at what seemed to be a Pareto front between the good representation of band structures and of elastic constants. This means further improvements in band structure resulted in worsening of the mechanical properties and vice versa. To verify that is the case, we compared hopping and overlap parameters that gave best band structures with a subset that resulted in the best mechanical properties. We used upper quantile planes from both sets and by taking the intersection approximated a region of parameter space that most likely will satisfy both requirements. Details of this approach will be published separately. This was an essential step to obtain the parameters presented in section 8. Having placed a new set of bounds on our model parameters we repeated the previous procedure with several further cycles of optimisation, this resulted in noticeable improvements. From the new set of results, we selected the most promising as a new starting point and changed the optimisation method. This time multi-stage local optimisation was applied. At each stage, we targeted only the bcc band structures for the majority spin with the Nelder–Mead method [32–34]. This way we had a much bigger chance of achieving a near-perfect fit and optimised the very subtle features of the band structure that are shaped by hopping integrals associated with p-orbitals.

In each stage of the optimisation, we selected a different search direction in the parameter space. For example, first, we optimised ratios for d-band parameters, then scales for each group of hopping integrals, together with the Stoner parameter and so on. Globally, we employed a semi-automatic framework where most of the steps were automated. However, the order and form of constraints would depend on the success of the optimisation. This way a sample of approximately 200 models was created. The three most promising candidates will be presented and evaluated in the following sections.

Except for additional tests presented in section 10, all calculations in this work used the Questaal software ([21]). In *ab initio* calculations we employed a full-potential augmented wave method with smoothed Hankel envelope functions. DFT calculations, due to similarities with the linear muffin-tin orbital (LMTO) method, and in an agreement with the Questaal website, will be referred to as full potential LMTO (FP-LMTO). User-specified basis parameters were converged with assumed precision of 10^{-6} Ry. This value was also set as the self-consistency energy threshold. Unless specified otherwise, the reference k-point density was no less than one represented by $32 \times 32 \times 32$ subdivisions of the primitive bcc cell. The exception was calculations of TB band structures during the optimisation of hopping parameters. In this case we used $8 \times 8 \times 8$ subdivision of the bcc primitive cell. While we are aware that this is not sufficient for precise total energy estimation, it was a fairly safe setting for estimation of band plots. It allowed us to greatly speed up the optimisation of hopping parameters.

8. Parameters of the electronic structure

For the sake of reproducibility, we now present details of electronic structure parametrisation. The fundamental relation between hopping magnitude and distance $|\overrightarrow{r}_{\alpha\beta}|$ between orbitals α and β, is given by an exponential decay i.e.

$$\tilde{h}_{\alpha\beta}\left(\overrightarrow{r}_{\alpha\beta}\right) = m_{\alpha\beta}\, \mathrm{e}^{-q_{\alpha\beta} r_{\alpha\beta}}, \tag{8}$$

where m controls the magnitude, q is the decay parameter and $r_{\alpha\beta} = |\overrightarrow{r}_{\alpha\beta}|$. The hopping function $h_{\alpha\beta}\left(\overrightarrow{r}_{\alpha\beta}\right)$ used in the calculation of Bloch sums and construction of the Hamiltonian

is simply

$$h_{\alpha\beta}\left(\overrightarrow{r}_{\alpha\beta}\right) = \tilde{h}_{\alpha\beta}\left(\overrightarrow{r}_{\alpha\beta}\right) \tag{9}$$

for $r_{\alpha\beta}$ smaller than the radius r_{A}. Otherwise,

$$h_{\alpha\beta}\left(\overrightarrow{r}_{\alpha\beta}\right) = \begin{cases} P_5\left(r_{\alpha\beta}, \tilde{h}_{\alpha\beta}, r_{\mathrm{A}}, r_{\mathrm{C}}\right) & \text{agm.} \\ \tilde{h}_{\alpha\beta}\left(\overrightarrow{r}_{\alpha\beta}\right)\, \tilde{P}_5\left(r_{\alpha\beta}, \tilde{h}_{\alpha\beta}, r_{\mathrm{A}}, r_{\mathrm{C}}\right) & \text{multi.} \end{cases}, \tag{10}$$

depending on the method of control of the tails of hopping integrals. The first case (agm) denotes augmentative cutoff where, $\tilde{h}_{\alpha\beta}$ is replaced by a fifth-order polynomial P_5 that matches value, slope and curvature of $\tilde{h}_{\alpha\beta}$ at r_{A}, while setting all three to zero at the critical radius r_{C} and beyond. In the second case of multiplicative cutoff (multi.) $\tilde{P}_5$ is a very similar polynomial with the exception of being normalised so it evaluates to 1 at r_{A}.

In the basis set we include $3d$, $4s$ and $4p$ orbitals. Furthermore, the Stoner parameter I was the same for all l-channels and the Hubbard-like U was set to 1Ry. In this regard, we follow the model of Paxton and Finnis [6].

The most promising models presented here are ones from the [QSGW, p_1, p_1] group. This means that the references were bands from QSGW calculations, the loss function was calculated using $p = 1$, the same measure was used to select the best sample from the optimised population. Each model is identified by a code from our data-set namely 73, 184 and 203. The resulting band structures are plotted against the reference and presented in figure 6. We focus on the ferromagnetic phase for all structures as this one was the reference. Note that in initial optimisation of all three band structures the Stoner parameter was much higher. It was reduced later to obtain near-perfect fit. It can immediately be seen from figure 6 that we obtained a very good agreement with the reference. Results for the fcc structure were calculated under strain as two TB models do not predict the meta-stable ferromagnetic state if volume per atom matches the stable bcc lattice.

Thanks to the *spd* basis we are able to recreate all the features up to energy levels that do not involve $4d$ orbitals. Although these levels are not shown here, in two cases we were able to maintain band topology in the case of the bcc structure (models 184 and 203). For this reason, these models have a shorter cut-off on the *sp* block of parameters. Needless to say, bands were tested under strain and they maintain most of the features, mainly for eigenvalues that do not involve p-orbitals.

Values of all parameters associated with the electronic structure are presented in table 3. The biggest difference between models is the decay parameters in the *sp* block as these were the parameters that were modified to obtain better mechanical properties in the last (semi-automatic) optimisation stage. However, in a set of over 200 models, no significant correlations (usually below 30%) between these quantities were found. Hence, at best, this step should be considered as a method to explore the subspace of parameters that provide a very good agreement in terms of band structures. From this subset, we can select ones that also allow us to find good elastic constants. In the future, we plan to find a way to find a relation between parameters within this region and optimal mechanical properties.

As mentioned before, the cut-off radius was modified in the case of models 184 and 203 to obtain a better representation of high-energy molecular orbitals in the bcc structure. Furthermore, there was a strong preference towards an orthogonal sp–d block. Therefore, in the case of model 184, this group of orbitals was explicitly set to be orthogonal. Apparent variations in

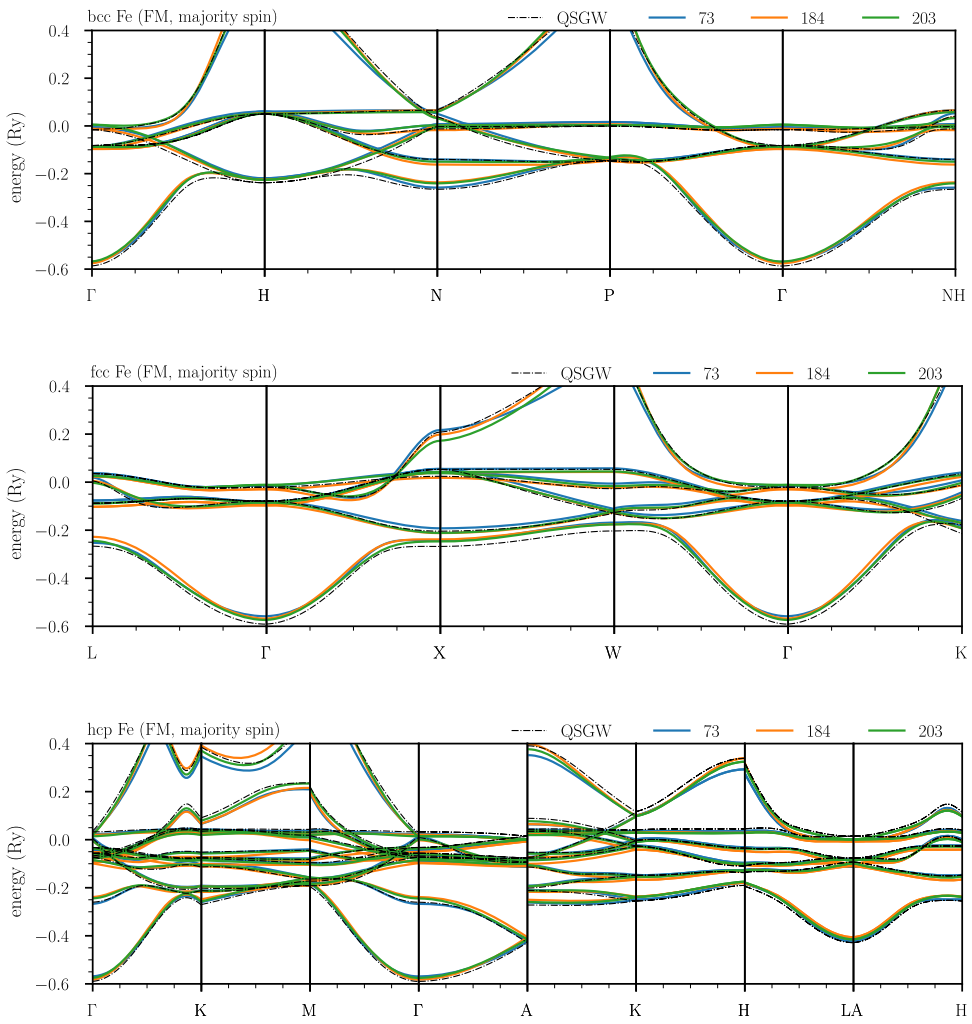

Figure 6. Band structures for bcc, fcc and hcp structures of Fe in the ferromagnetic state. Here, the fcc lattice is under expansion 2% and .35% for models 73 and 184 respectively. The black dotted line represents band structures plotted from electron density updated in the QSGW self-consistency cycle.

other parameters might be misleading due to ambiguity arising from accepting hopping magnitude as a function of bond length. This aspect of the models is illustrated by a logarithmic relationship between magnitude and decay (see e.g. figure 7).

It is clear that all estimates of optimal hopping integrals that involve *d* orbitals are very similar, although these differences are sufficient to influence the mechanical properties. Not surprisingly, the biggest variations can be observed in the case of *sp* orbitals.

9. Elastic properties

The models presented in this paper are based on two types of repulsive pair-potentials—V. The first is referred to as Goodwin–Skinner–Pettifor (GSP, [35]) and has the form

Table 3. Table of TB parameters of Fe associated with the electronic structure. The Hubbard-like U is 1Ry in all cases. Orbitals included in the model are $3d$, $4s$ and $4p$. Cut-off radii $r_{\mathrm{A}/\mathrm{C}i}$, where $i = 1, 2, 3$, correspond to following hopping integrals: $1-\{dd\sigma, dd\pi, dd\delta\}$, $2-\{sd\sigma, pd\sigma, pd\pi\}$, $3-\{ss\sigma, sp\sigma, pp\sigma, pp\pi\}$.

Off-diagonal hopping parameters

	Hopping						Overlap					
	Magnitude (Ry)			Decay (a_0^{-1})			Magnitude (1)			Decay (a_0^{-1})		
	73	184	203	73	184	203	73	184	203	73	184	203
$ss\sigma$	−1.935	−1.350	−3.740	0.600	0.500	0.742	1.984	1.160	2.500	0.600	0.500	0.742
$sp\sigma$	1.927	1.169	3.974	0.600	0.500	0.742	−2.200	−1.142	−3.500	0.600	0.500	0.742
$pp\sigma$	2.100	1.600	4.730	0.600	0.500	0.742	−2.809	−0.950	−4.200	0.600	0.500	0.742
$pp\pi$	−1.048	−0.500	−2.090	0.600	0.500	0.742	0.350	0.143	0.150	0.600	0.500	0.742
$sd\sigma$	−2.711	−3.349	−1.783	0.900	0.950	0.808	0.000	0.000	0.636	0.000	0.000	0.808
$pd\sigma$	−2.232	−3.104	−1.492	0.900	0.950	0.808	0.000	0.000	0.710	0.000	0.000	0.808
$pd\pi$	2.179	2.754	1.492	0.900	0.950	0.808	0.000	0.000	−0.658	0.000	0.000	0.808
$dd\sigma$	−2.322	−3.186	−4.874	0.900	0.950	1.058	0.000	0.000	0.000	0.000	0.000	0.000
$dd\pi$	1.633	1.922	3.555	0.900	0.950	1.058	0.000	0.000	0.000	0.000	0.000	0.000
$dd\delta$	−0.499	−0.386	−0.917	0.900	0.950	1.058	0.000	0.000	0.000	0.000	0.000	0.000

On-site, cut-offs and Stoner parameter

	ϵ_{3d}(Ry)	ϵ_{4s}(Ry)	ϵ_{4p}(Ry)	$r_{\mathrm{A}1}(a_0)$	$r_{\mathrm{A}2}(a_0)$	$r_{\mathrm{A}3}(a_0)$	$r_{\mathrm{C}1}(a_0)$	$r_{\mathrm{C}2}(a_0)$	$r_{\mathrm{C}3}(a_0)$	I
73	−0.002	0.248	0.718	8.504	8.504	8.504	10.087	10.087	10.087	0.046
184	−0.004	0.346	0.843	8.504	8.504	6.525	10.087	10.087	8.504	0.049
203	0.000	0.350	0.830	8.504	8.504	6.525	10.087	10.087	8.504	0.049

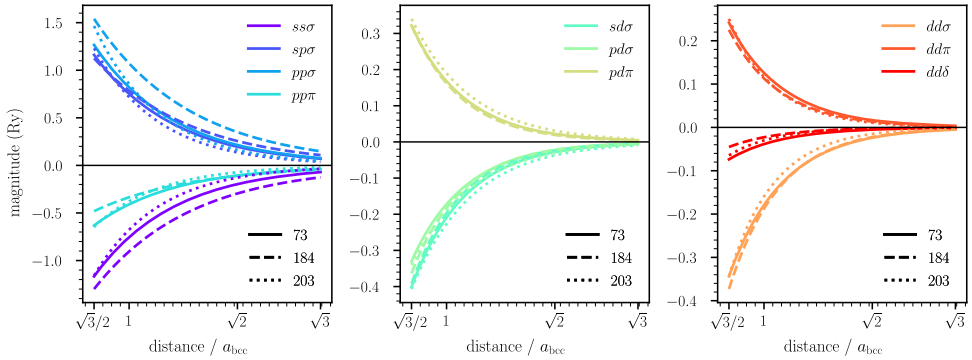

Figure 7. Nominal values (excluding augmentation) of hopping integrals as a function of distance between orbitals. Ticks indicate subsequent neighbour shells in the bcc structure.

$$V_{\mathrm{GSP}}\left(\overrightarrow{r}\right) = A\left(\frac{r_0}{r}\right)^n \exp\left(-n\left(\frac{r}{r_\mathrm{c}}\right)^{n_\mathrm{c}} + n\left(\frac{r_0}{r_\mathrm{c}}\right)^{n_\mathrm{c}}\right), \tag{11}$$

where $r = |\overrightarrow{r}|$ and n, n_c, V, r_0 and r_c are adjustable parameters, each controlling different features of the potential. Here, we apply a shortened notation where dependence on the parameters is implicit. The second type of potential has a similar form to that for the hopping integrals i.e.

$$V_{\mathrm{exp}}\left(\overrightarrow{r}\right) = \left(\sum_{i=1}^{3} a_i r^{b_i} \exp\left(-c_i r\right)\right), \tag{12}$$

where for evaluation of values of the potential we use the same augmentation procedure as in the case of hopping integrals and overlap functions (equation (10)). Hence,

$$V\left(\overrightarrow{r}\right) = V_{\mathrm{exp/GSP}}\left(\overrightarrow{r}\right)\tilde{P}_5\left(r, V_{\mathrm{exp/GSP}}, r_\mathrm{A}, r_\mathrm{C}\right). \tag{13}$$

Here we distinguish between r_c and r_C. The former is an internal parameter of V_{GSP} function while the latter is a global cut-off parameter that defines which interactions will be set explicitly to zero. Potential V_{GSP} does not need to be augmented as it does not have such heavy tails as V_{exp}. Nonetheless, the code was set to augment this potential right before r_C. Furthermore, the value of r_C was set to $2a_{\mathrm{bcc}}$, where a_{bcc} is the equilibrium-volume lattice parameter. When using V_{exp} the augmentation becomes a critical factor and both associated parameters, r_A and r_C, were also subject to optimisation.

In our experiments we preferred V_{exp} due to its simplicity. However, after testing several other forms of pair-potentials available in the Questaal code, we found that we were able to improve some results using V_{GSP}.

In optimisation of the pair-potentials, we could not rely on QSGW calculations as a reference. The main reason is that we required a dense sampling of the total energy as a function of the strain tensor, and at the moment, these calculations were too expensive. As mentioned above, we instead used FP-LMTO with the PBE91 exchange–correlation functional [21, 36]. The objective was to recreate energy–strain curves in deformations defined in table 1. The results are presented in figure 8.

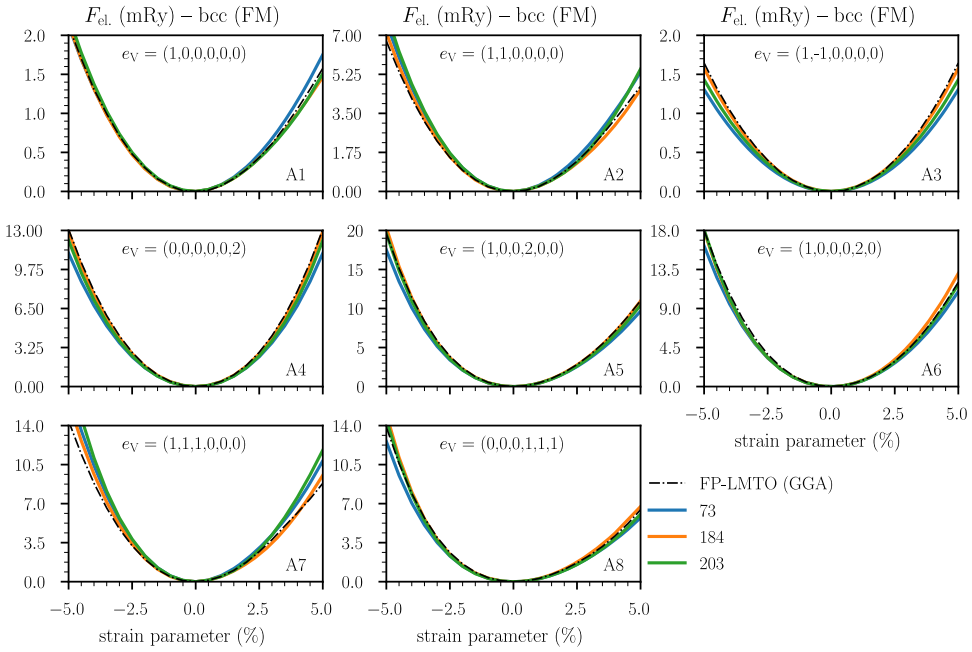

Figure 8. Energy–strain curves for eight different strain patterns. Energy is normalised by the number of atoms in a unit cell. Strain patterns are given in Voigt notation. Piecewise linear interpolation was used and the distance between data-points is 0.5%.

Table 4. Parameters of pair-potentials. Parameters not specified in lower table are set to 0. Note that Questaal code uses Atomic Rydberg Units.

Model with V_{GSP} pair-potential

	r_A	r_C	A	n	n_c	r_0	r_c
73	10.757	10.811	931.102	7.481	10.567	1.312	6.337

Models with $V_{exp.}$ pair-potential

	r_A	r_C	a_1	b_1	c_1
184	4.353	6.493	1399.534	−1.037	1.817
203	4.379	6.832	768.104	−0.000	2.004

Estimates of optimal parameters are given in table 4. It immediately transpires that we were able to obtain a good agreement with the reference. The tendency to underestimate total energy in case of A3 is preferable since the DFT tends to overestimate the c' elastic constant.

In the calculation of elastic constants, instead of the classical approach i.e. fitting a polynomial to an energy–strain curve and taking an appropriate derivative, we fit Brugger's equation of state (EOS) ([37]) to all relationships at the same time. We use the explicit form for cubic materials as given by Vekilov *et al* in [26]. By doing so, we reduce the dependency of our results on the selection of deformation patterns and order of calculations. Note that, in principle, it is

possible to calculate second-order elastic constants (SOEC) from at least three deformation patterns. There is an infinite number of patterns (although they are not arbitrary) we can use and results may vary significantly. We decided to use an equation-of-state that takes into account up to third-order constants (TOEC). For that reason, we used results of calculations of elastic constants for strain magnitudes no greater than 0.025. Otherwise, it was impossible to recreate energy–strain curves as it would require a more flexible (higher-order) EOS. Also note that this meta-parameter is another way that allows us to adjust the results. Therefore, one needs to be careful when interpreting them. While SOEC are not very sensitive to such manipulations, some TOEC can be greatly affected by the selected strain magnitude.

It is fairly simple to assess SOEC since high-quality experimental data are available. Here, we relied on the more recent publication by Adams *et al* ([38]) instead of the classical paper by Rayne ([39]). However, it is much more difficult to obtain results for TOEC at temperatures near 0 K. For that reason, we used a mixture of theoretical and experimental results in the reference data set, where experimental results include also measurements at the room temperatures.

Furthermore, we assess predictions by our TB probabilistically. For each quantity of interest (QOI) and each model, we estimate expectation and variance. The test statistic is inspired by the *t*-test and will be given by

$$t_{\alpha\beta} = \frac{X_{\alpha\beta} - m_\beta}{s_\beta}, \tag{14}$$

where X is the matrix of TB results, α indexes the model while β the QOI, m corresponds to test-set mean and s sample standard deviation. To interpret data we can compare values of t to quantiles of the normal distribution with zero expectation and unit variance—$\mathcal{N}(0, 1)$. The most important quantiles are $0.975/0.025$ that correspond to values ± 1.96 of $t_{\alpha\beta}$.

Results are presented in table 5. It is clear that the model 184 provides an excellent agreement with the reference data with respect to the SOEC. However, estimates of TOEC are far from perfect. Although given the data we can test only the consistency with the set. The set itself is rather arbitrary as it is based on the availability of the data. Furthermore, uncertainty associated with the data is unknown.

10. Transferability

The most important aspect of a model is its transferability. In this work, we are most concerned that our model will be able to predict energies and associated forces in bcc structures with defects. However, the predictive value can be assessed by the estimation of formation energies and elastic constants for other structures. In principle, the family of parametric TB models has a significant advantage over analytical or purely numerical 'surrogate' models. Although the definition of the pair-wise repulsion is arbitrary to some degree, a big part of the model is a valid approximation to the many-body problem. In this section, we investigate the behaviour of our models in non-native structures.

The first consideration are formation energies ($E_{\text{form.}}$) with respect to the bcc structure in the ferromagnetic state (FM) i.e.

$$E_{\text{form.}}\,(\text{struc.}) = \min_V \left\{ E\left(V|\text{struc.}\right) - E\left(V_0|\text{bcc}\right) \right\}, \tag{15}$$

where the function min returns the minimum of a set, and struc. refers to the crystallographic structure in question. Minimisation is carried out over volume per atom V and V_0 corresponds to the ground state of the bcc (FM) structure at 0 K. In practice, we evaluated energies over

Table 5. Estimates of elastic constants given in GPa. References are as follows: 1—experimental data, [40, 41] as quoted in [42], 2—experimental data [38], 3—DFT (GGA) estimate, [42], 4—experimental data, high temperatures, [43], 5—DFT (GGA) estimate, this work.

Elastic constants in (GPa)											
	073		184		203		Reference				
	val.	t	val.	t	val.	t	1	2	3	4	5
c_{11}	288.41	1.60	269.80	0.80	286.25	1.51	243.00	239.55	263.04	226.00	285.21
c_{12}	192.57	4.34	154.31	0.83	179.69	3.16	138.00	135.75	151.14	140.00	161.61
c_{44}	95.67	−2.56	104.99	−1.32	97.97	−2.26	122.00	120.75	103.34	116.00	112.51
c_{111}	−3038.71	0.95	−4277.47	−0.22	−5954.21	−1.81	−4820.00	—	−4971.50	−2720.00	−3665.98
c_{112}	−1195.87	−13.45	−1367.62	−17.86	−1067.33	−10.19	−700.00	—	−675.14	−608.00	−673.96
c_{123}	−975.66	−0.63	−1349.60	−0.86	−2124.49	−1.35	2460.00	—	−806.29	−578.00	−891.74
c_{144}	−1014.91	0.17	−1047.92	0.09	−930.28	0.39	−1580.00	—	−699.01	−836.00	−1220.63
c_{155}	−801.47	0.01	−557.97	0.90	−655.48	0.54	−1030.00	—	−606.69	−530.00	−1046.25
c_{456}	−697.73	−0.39	−634.81	−0.27	−743.97	−0.48	275.00	—	−647.01	−720.00	−880.67
c'	47.92	−0.75	57.74	0.69	53.28	0.04	52.50	51.90	55.95	43.00	61.80
B	224.52	2.99	192.81	0.83	215.21	2.36	173.00	170.35	188.44	168.67	202.81

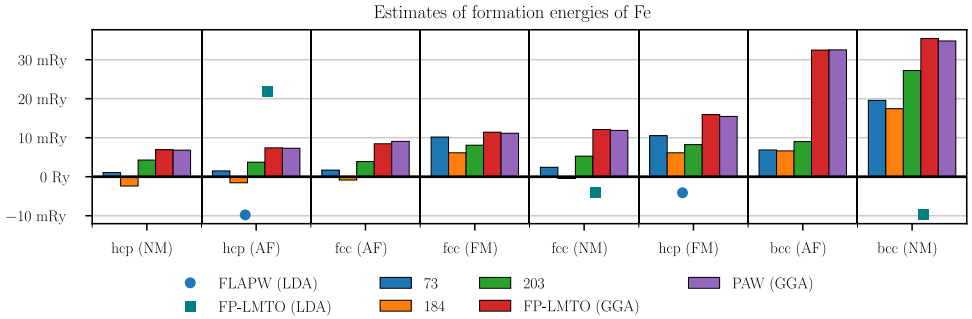

Figure 9. Formation energies of Fe, relative to the binding energy of the bcc (FM) phase, estimated using variety of models. Results for FLAPW (LDA) are taken from [46]. LDA results are included to illustrate challenges associated with approximation of exchange and correlation effects. However, they should not be considered as the optimality limit of this method.

the range from -0.2 to 0.2 of engineering volumetric strains with a 0.005 increment. Note that not all calculations completed successfully. For example, calculations with the ferromagnetic starting point were likely not to converge near strains that favour the non-magnetic state. Results were interpolated by fitting a 12th order polynomial

$$E(V|\text{struc.}) = \sum_{n=0}^{12} \alpha_n \left(\frac{1}{V^{1/3}} \right)^n . \tag{16}$$

In the process we used the standard least-square methods. The form of the function was inspired by the SJEOS equation of state [44] with some extra degrees of freedom added to improve the fit. In this case we do not require high numerical precision and any reasonable interpolator will suffice.

The focus is on bcc, fcc and hcp structures in ferromagnetic, non-magnetic (NM) and antiferromagnetic (AFM) phases. We consider non-co-linear magnetism with a single-layer on an arbitrary closed-packed direction in the AFM case. Such spin configuration does not minimise the total energy although should be sufficient for comparison.

We begin with analysis of a wide range of methods. The data presented in figure 9 consists of the following: results from TB calculations—model 73, 184, 203; full-potential LMTO method implemented in Questaal ([21]) with GGA (PBE) and LDA functionals (FP-LMTO (GGA) and FP-LMTO (GGA) respectively); PAW (projector-augmented-wave) method with GGA implemented in VASP [45]. In the latter case we used 600 eV cut-off energy and $24 \times 24 \times 24k -$ point grid on a bcc unit cell. In summary, we have the following categories of results: TB, DFT with different exchange–correlation functionals and DFT with two types of basis-set. Before addressing TB results we consider DFT results first. Calculations in Questaal and VASP were made by the present authors and these should be considered as the main point of reference. The data from the literature shows results from the development of pseudopotentials. The main reason for including them, as well as results from LDA calculations, is the management of expectations and validation of primary references.

It immediately transpires that Questaal and VASP results are almost identical and the basis set, if optimised, does not influence results. Furthermore, the LDA results show negative formation energies for non-magnetic close-packed structures. This is the most important result in the context of what can be expected from TB models. In other words, we cannot expect exact

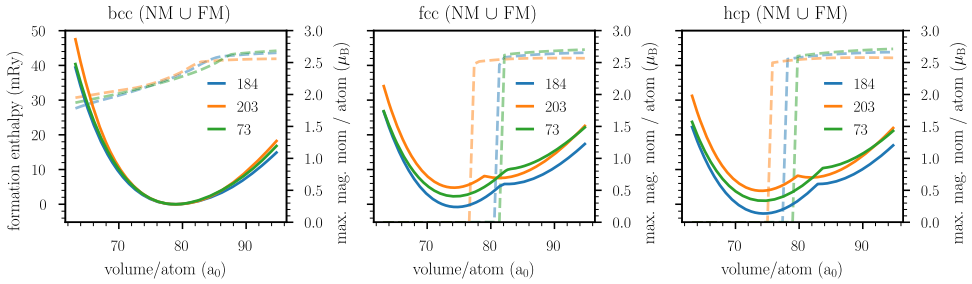

Figure 10. The lower value from two total energy–volume curves (FM and NM phases only), relative to the bcc formation energy. The curves represent data with piece-wise linear interpolation. The dashed lines are plotted against the right axis and correspond to magnetisation per atom. Note that the total energy is shifted by the reference energy which depends on the initial charge configuration. The purpose of these graphs is to illustrate correctly predicted (in a qualitative sense) transition from one magnetic state to another under the volumetric strain. For the SCTB model to converge to a magnetic state calculations need to be initiated with broken symmetry between spin channels. To demonstrate a spontaneous transition from one magnetic state to another, we plot magnetisation for calculations initiated in such a way. Hence, kinks in energy–volume curves (a minimum from two sets of results: magnetic and non-magnetic) do not exactly correspond to these spontaneous changes.

estimates from TB when *ab-initio* methods are predicting different than bcc (FM) ground states of Fe at 0 K. This is indeed the case in LDA calculations. From comparison with the literature, we also know that is not a result of an error on our part.

Figure 9 provides quantitative assessment of TB predictions. As expected, we can hope for qualitative agreement at best and only with respect to the ordering of non-equilibrium phases. The only exception here is the fcc (FM) case, where formation energy is greater than the overall trend would suggest. However, note that the initial charge distribution affects reference energy in TB calculations. This makes it difficult to assess the results. Nonetheless, the results are directly comparable within a given magnetic phase.

Next, we consider the transition from the magnetic state under pressure. If magnetism is correctly included in the model, such transformation should occur in closed-packed structures. Predictions of the TB models are illustrated in figure 10. In a broader sense, our models behave as they should. This is consistent with DFT results. Namely, within the considered range of strains, the FM state stabilises the bcc structure. Close-packed structures will experience a transition from the FM to NM state under compression. All models predict similar transition points which is consistent with DFT results. The above also illustrates that in some cases there is no minimum in the metastable ferromagnetic state. Simply, for these data points, convergence criteria were not met. Such issues can be avoided by adjusting charge convergence parameters. However, we decided to use universal settings for all calculations. Hence, missing data were filled by extrapolation of a 3rd-order polynomial fitted to several last available data points.

In the context of transferability into structures with defects, mechanical properties in non-equilibrium structures might provide a better indication of performance. We limit our calculations to estimates of elastic free energy and calculations of the bulk modulus. While this is not a sufficient indicator it is a necessary one. The results are presented in figure 11. The performance varies significantly depending on the structure and magnetic phase. Differences in estimates for the bcc (FM) case are quite small as this was included in the loss function. We

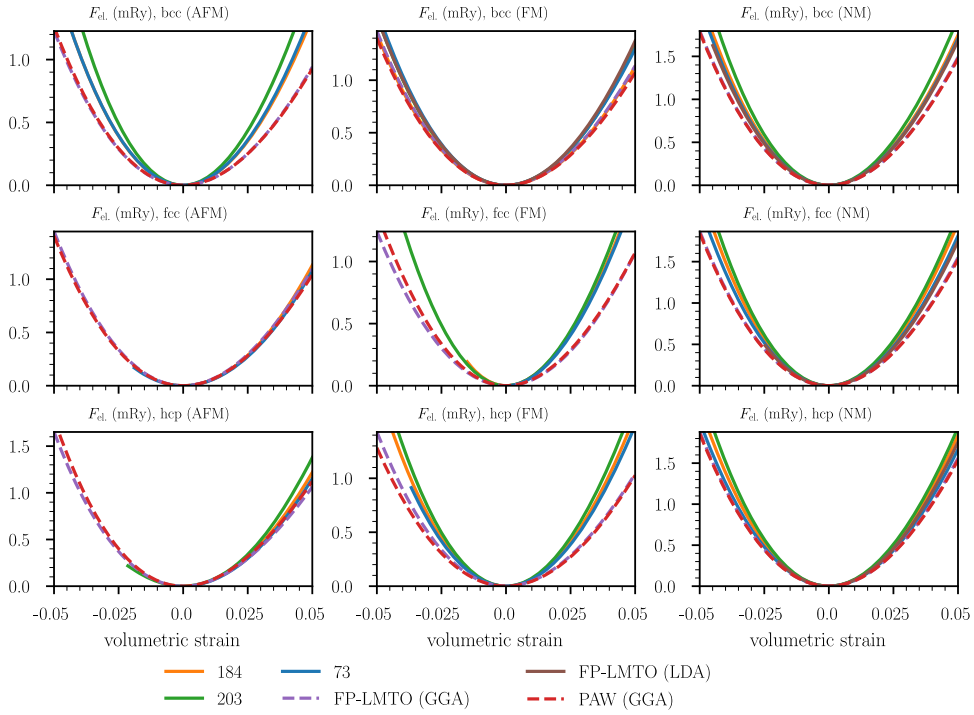

Figure 11. Estimates of elastic free energy per atom as a function of the volumetric strain in Fe. Each row represents a given structure and each column a magnetic state.

can see that TB and DFT (including GGA and LDA) are in a very good agreement. Furthermore, we can see a fair agreement in non-magnetic phases and mostly very good agreement in anti-ferromagnetic phases.

We emphasise that figure 11 shows elastic free energy per atom versus volumetric strain. Estimates of elastic energy density would be different since each method predicts a different equilibrium volume. While TB was forced to fit the experimental lattice parameter (2.866 Å) of the bcc (FM) structure, DFT with GGA functional would slightly underestimate it. Predictions of the DFT method within the LDA resulted in significant underestimations. In some cases, this method failed to find a minimum even within the range of ± 0.2 strain (with the experimental lattice parameter for bcc Fe as a reference point). This reflects on better estimates of bulk modulus in case of the TB method and much worse performance of the LDA than figure 11 would suggest. This is clearly illustrated in table 6 and figure 12.

In summary, the performance of the TB method in this test is no worse than LDA. Except for the bcc (FM) case, the earliest model no. 73 seems to perform better than its counterparts. We can also see that these estimates differ from ones presented in table 5. In this test, we simply report the normalised value of the second derivative at equilibrium instead of fitting a complete model of the elastic energy as a function of the strain. Therefore, a slightly worse performance of the model 203 needs to be considered with some extra care.

Nonetheless, we focus here on the large deviations from the reference. We assume that the problem arises from magnetic interactions for phases that are far from the equilibrium. For example, a test on fcc (AFM) demonstrated a fairly good agreement with DFT (GGA) calculations. On the other hand, fcc (FM) performed quite poorly. According to predictions by

Table 6. Estimates of bulk modulus of Fe at 0 K in GPa. In the case of hcp structures c/a ratio was kept constant.

	bcc (AFM)	bcc (FM)	bcc (NM)	fcc (AFM)	fcc (FM)	fcc (NM)	hcp (AFM)	hcp (FM)	hcp (NM)
184	225.0	187.9	309.5	175.7	273.4	341.5	200.9	243.4	318.0
203	278.0	213.0	342.5	—	268.3	366.8	200.2	272.7	342.4
73	225.9	219.1	294.4	174.1	248.2	313.6	190.2	228.9	288.5
FP-LMTO (GGA)	172.7	201.8	272.6	209.0	169.5	289.7	227.2	180.8	293.7
FP-LMTO (LDA)	—	247.4	328.1	—	—	346.3	—	—	351.5
PAW (GGA)	174.2	189.5	269.4	200.3	182.6	285.4	244.8	166.7	289.2

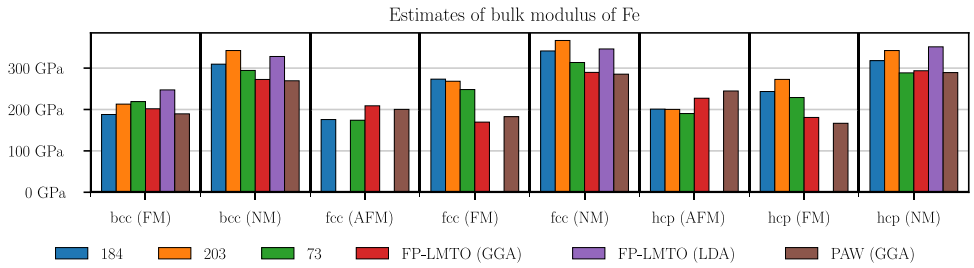

Figure 12. Estimates of the bulk modulus. Some test-cases were omitted due to lack of well a defined minimum.

Wróbel *et al* ([47]) the ground state of fcc phase (at 0 K) corresponds to the anti-ferromagnetic double-layer (AFMDL) spin configuration. It seems that the pattern is as follows: the further away we are from the ground state—in terms of magnetic configuration, the less consistent with the DFT (GGA) are the estimates of elastic free energy. This effect includes calculations for fcc and hcp structures. Although we can only speculate at this point, these results have a reasonable explanation. The non-selfconsistent Hamiltonian (the starting point of the TB calculations) at best corresponds to a variational approximation to the ground state density. Implicitly, by setting parameters, we specify a starting point on which a variational approximation depends. In such context, it seems that there is a limited energy window in which TB models can be transferable. On the other hand, it is also possible that by prioritising a good fit of the band structures we underestimated the overlaps and lost transferability of the hopping integrals through inaccurate consideration of the environmental dependence (see e.g. Paxton *et al* in [10, 48]).

11. Defects

In the scope of this paper, the most important tests of transferability are the self-interstitial and mono-vacancy formation energies ($E^{\mathrm{f}}_{\mathrm{i(v)}}$). This relatively small test set will provide a measure of the predictive value of the model. These energies are given by

$$E^{\mathrm{f}}_{\mathrm{i(v)}} = E_{\mathrm{i(v)}} - \left(\frac{N \pm 1}{N}\right) E_{\mathrm{bulk}}, \tag{17}$$

Table 7. Defect formation energies, given in electronvolts, where ΔE^{f} is the difference between estimates for the $\langle 110 \rangle$ dumbbell and a monovacancy. The following references were used: ()[1]—[49], ()[2]—[9], ()[3] and ()[4]—multiple sources in [49]. The acronym GAP corresponds to Gaussian approximation potential that uses machine-learning methods to estimate an optimal classical potential surrogate [50].

Model	$E^{\mathrm{f}}_{\mathrm{v}}$	$E^{\mathrm{f}}_{\mathrm{i}\,\langle 110 \rangle}$	ΔE^{f}
73	1.87	2.72	0.85
184	2.60	3.60	1.00
203	2.03	2.85	0.82
GAP[1]	2.26	4.21	1.95
TB (d)[2]	1.75	4.36	2.61
DFT[3]	2.07–2.15	3.64–4.02	—
Experiment[4]	0.55–2.00	4.70–5.00	—

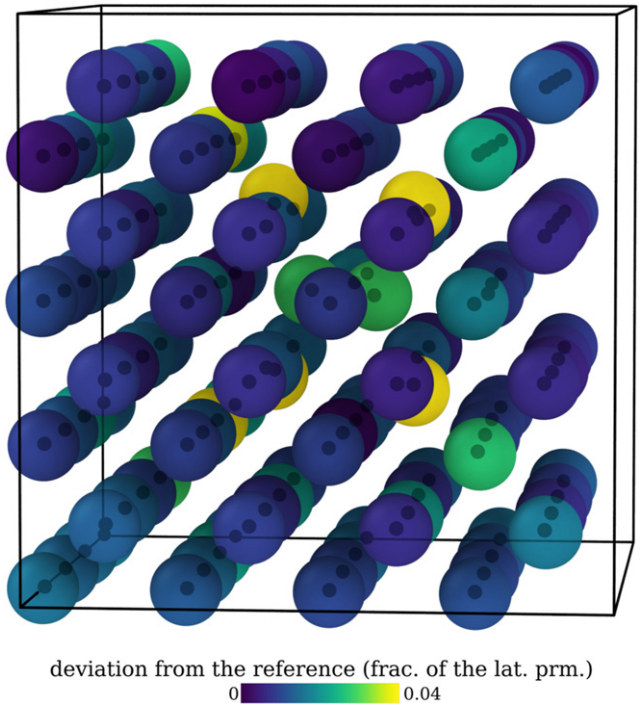

Figure 13. Difference between dumbell structures as estimated by the DFT and TB (model 184). The difference is given in fractions of the equilibrium lattice parameter for the bcc Fe (FM) as predicted by the given method. Black circles are representing DFT estimates and spheres the TB. Colour corresponds to the norm of differences between relative positions. In case of the vacancy, structure deviations are no greater than 2% of the corresponding lattice parameter.

where E_{bulk} is the total energy of a computational cell, containing N atoms, without the defect, while E_{i} and E_{v} are the total energy of a cell containing a self-interstitial or a vacancy respectively. Estimates are made for a variety of volumes with fully relaxed atomic positions. The results are presented in figure 14 and summarised in table 7. Figure 13 shows comparison of

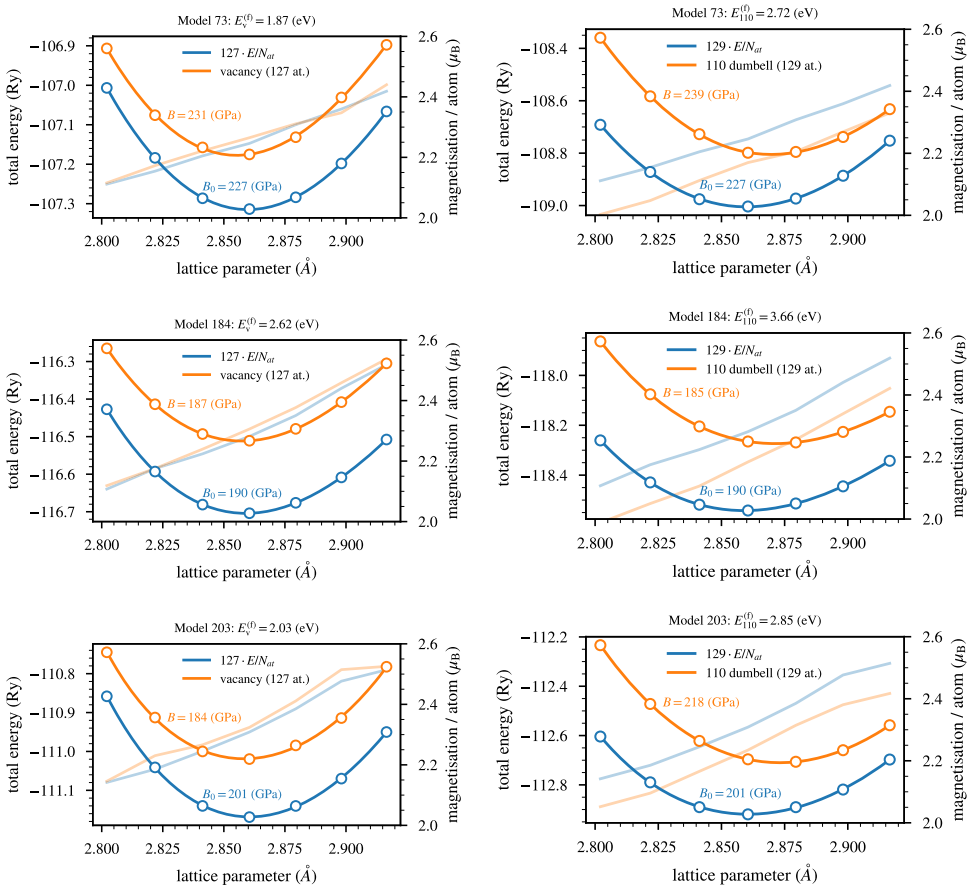

Figure 14. Formation energy (left axes) and magnetisation (right axes) of defects in the pure bcc (FM) iron. The bulk modulus for the reference energy–volume curve (blue lines) is calculated before scaling. In calculations, we used 4th order polynomial and evaluated 2nd derivative at the minimum. Calculations were made with the initial $4 \times 4 \times 4$ number of k-points in a $(4 \times 4 \times 4) \times a_{\mathrm{bcc}}^3$ computational cell. Such sampling usually results in an extra 2% numerical uncertainty associated with estimates of the formation energies.

the interstitial structure as predicted by DFT (PAW with GGA) and TB. The difference will be given by the Euclidean (L_2) norm

$$\left\| \vec{x}_{\mathrm{DFT}}/a_{0(\mathrm{DFT})} - \vec{x}_{\mathrm{TB}}/a_{0(\mathrm{TB})} \right\|_2, \tag{18}$$

where $a_{0(\mathrm{DFT/TB})}$ is the bcc lattice parameter predicted by the corresponding method, while $\vec{x}_{\mathrm{DFT/TB}}$ is the position of an atom.

We begin with the consideration of three major characteristics of the energy–volume curves (figure 14): lattice parameter, magnetisation and bulk modulus. Unlike the self-interstitial, introduction of a vacancy has no major effect on the first two quantities. As one might expect, the introduction of an extra atom reduces the magnetisation per atom and increases the equilibrium lattice parameter. Values of the bulk modulus decreased (vacancy) or remained largely unchanged (interstitial). Qualitatively, these changes can be considered as reasonable.

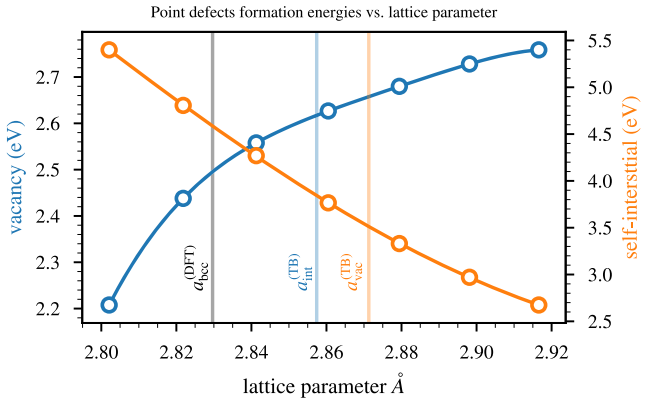

Figure 15. Comparison of formation energies of a vacancy and self-interstitial (model 184) as a function of the lattice parameter.

However, estimates of elastic constants are extremely sensitive. Most likely this is a numerical artefact arising from the insufficient number of k-points. It would appear that models 73 and 203 provided a very good estimate of the vacancy formation energy while failing in the case of an interstitial defect. On the other hand, the difference between the lowest DFT estimate of the interstitial formation energy and worst TB prediction (model 73) is below 25%. Model 184 is consistent with the DFT predictions of interstitial formation energy while giving a bit too high estimates in case of a vacancy (at least 30% higher than DFT). Nonetheless, differences between formation energies suggest that this model is performing better.

So far it seems that by targeting elastic constants we can improve our predictions of the defect formation energies, although other factors might be important as well. For example, TB models were optimised towards 2.2 μ_{B} magnetisation per atom in the perfect bcc cell.

It is possible that this resulted in the underestimation of the magnetic contributions and we should sacrifice this property. However, in the optimisation of the elastic constants, higher values of the Stoner parameter, that controls the magnetisation, resulted in undesirable changes in energy–strain curves. It is also possible that we need to manage our expectations and our estimates are as good as they can get without over-fitting (sacrificing predictive value for the sake of improvement of known quantities) or the introduction of cancelling errors. In terms of structures of the defects, we obtained a very good agreement with DFT results, as illustrated in figure 13. Furthermore, when one considers formation energies for a fixed lattice parameter (figure 15), it is apparent that results are qualitatively correct for bond lengths predicted by the DFT method.

Finally, in the case of a vacancy defect, we are quite close to GAP calculations that contained a vacancy in the fitting set. These findings are likely to be relevant, given that mechanical properties are based on energy–strain curves calculated using this method. In other words, we might need a better than DFT method to obtain a reference for optimisation of the mechanical properties.

12. Summary and conclusions

In this paper, we investigated a method of obtaining parameters for the self-consistent polarisable TB model of Fe, to study diffusion in low-alloy steels in the future. We decided to use

the non-orthogonal *spd* basis which means that 3*d*, 4*s* and 4*p* atomic orbitals are implicitly included in the construction of the initial Hamiltonian with the SK algorithm. As a result, it was possible to recreate most features of the band plots for bcc, fcc and hcp structures. Additionally, as band calculations are fairly inexpensive, it was possible to optimise the model with a minimum number of constraints. This led to another advantage of our approach i.e. the ability to include in the objective data from methods more 'complete' than the DFT, such as QSGW.

Furthermore, we introduced a more robust optimisation of pair-potentials that takes into account higher-order terms. As demonstrated in figure 5, estimation of second-order elastic constants is insufficient to confirm the correctness of the pair-potentials. This way we achieved a better sampling of pair-potentials that is more sensitive to pairs outside the 1st neighbour shell.

Finally, we were able to address some interesting aspects of how to design an optimisation strategy. This includes a selection of the loss function, optimisation algorithms and a strategy. We found that the covariance matrix adaptation strategy (CMAES) is a good algorithm to optimise a large set of TB parameters, while the simplicial homology global optimisation (SHGO) is superior in the optimisation of pair-potentials. Note that the SHGO algorithm is designed for a small number of dimensions (<10), hence it could not be used in the optimisation of band structures.

The novelty of our approach can be regarded as putting more emphasis on maximisation of the predictive value rather than minimisation of the loss function over a rich test-set. This means we decided to optimise parameters on a small, computationally inexpensive set while iteratively focussing the search region near the most promising candidates, evaluated in subsequent stages.

Our approach was tested in a series of numerical tests aimed at a recreation of key properties of allotropes of Fe. We were able to achieve a very good agreement with reference band plots, generated using the QSGW method. It is likely that this was possible thanks to a limited number of constraints and the full *spd* basis. Surprisingly, out of three tested references (DFT (LDA), DFT (GGA) and QSGW), QSGW-based band structures also resulted in the best representation of the elastic properties. Although we achieved a good agreement with experimental second-order elastic constants, representation of third-order elastic constants can be regarded as qualitative at best. In the context of transferability, the results are consistent with DFT (GGA) results assuming that the formation energy of a given phase does not deviate too much from the bcc (FM) binding energy.

While energy–volume curves for non-magnetic phases are in a good agreement with the reference, clearly results associated with non-native ferromagnetic phases could be improved. The problem can be traced to putting too much emphasis on the representation of the band structures. It is possible that we explored a basin of orthogonal models that can have a limited transferability. Figure 7 illustrates how difficult is to separate both cases. The data suggest that magnetism could be responsible for the under-performance and adjustments of the Stoner and Hubbard-like U parameter, might improve the results.

On the other hand, the performance of our TB model is comparable to other models. For example, Bacalis *et al* underestimated the bulk modulus of the ferromagnetic bcc Fe ([8]) as did Madsen *et al* ([11]). However, in the latter paper, we can find very good results with respect to the bulk modulus in other phases. Liu *et al* ([9]) achieved overall very good values of bulk modulus, except for ferromagnetic and anti-ferromagnetic hcp phase. Additionally, if we consider the model by Paxton and Elsässer [10], and also consider estimates of bulk modulus as a simplified measure of transferability, it seems that this aspect is universally problematic. Needless to say, a single elastic constant is not a good performance measure although deviations from the reference value suffice to indicate the level of the transferability.

In the context of our research, the most important test was the estimation of formation energies of the self-interstitial and the vacancy. Predicted structures of point defects are in good agreement with the DFT results. While it is a less robust test than e.g. evaluation of a dislocation structure, it illustrates that a model (as a set of parameters) generated using our method can be used as a low-fidelity model to speed-up relaxation of atomic positions while assisting DFT calculations. Given the influence of magnetism, this result can be considered as a positive indication.

With respect to formation energies, when comparing our results with the whole range of those generated with the DFT method, the performance of our models could be viewed as satisfactory. However, it is far from the experimental values. Furthermore, the d-band TB model by Liu *et al* ([9]) performed much better as well as the GAP developed by Dragoni *et al* ([49]) even though magnetism cannot be treated appropriately with purely empirical models.

In defence of our model, we can point out that these results were not part of the fitting data-set and should be regarded as a pure prediction. Therefore, 20%–30% deviations from DFT predictions is something that one might expect. Having said that, the utility that our model can provide might be limited. The solute–vacancy binding energies in iron-based alloys are usually of the order of magnitude of 0.1 eV (see e.g. Messina *et al* [51] or Whiting *et al* [52]). Although, in the context of potential applications of the TB method, more interesting are estimates of potential barrier in simulations of defect migration. In this associated errors might be at the same order of magnitude as quantities of interest.

We speculate that this problem is partially associated with forcing TB models to stabilise the equilibrium lattice parameter (figure 15) (as well as with difficulty of optimising overlaps together with hopping integrals). Self-consistent TB is consistent with the local spin-density approximation which tends to heavily underestimate the bond length of Fe. By including *s* and *p* orbitals we introduce an extra level of ambiguity as well as add interactions that cannot be well represented in this method (e.g. hopping between *p* and *d* atomic orbitals).

We were not able to develop a model that is superior in terms of transferability or mechanical properties. However, we found a way to generate models with comparable performance. Therefore, we provide an alternative fitting route for the semi-empirical SCTB. We view our approach as complementary to existing methods. We also believe there are applications where our method could prove to be more suitable. As mentioned throughout this paper, SCTB models with a complete *spd* basis pose a particularly challenging optimisation problem. This can be partially overcome by taking advantage of deep knowledge and intuition or by applying sophisticated techniques, such as ones that derive parametrisations from first principles. However, such expertise might not be available.

In material science, researchers are often involved in developing multi-scale multi-physics frameworks that must fit well into the existing software infrastructure. In such a context, mastering a sophisticated code or developing a new one is a considerable investment, making an agnostic data-driven approach a compelling alternative. Firstly, it relies on common and transferable knowledge of mathematical optimisation. Secondly, it can be applied using any TB implementation. In our work, we were using a set of bash and Python scripts to call externally the Questaal software package ([21]). This package can be easily replaced by any other code.

The data-driven approach provides other benefits as well. Although, the following argument does not apply to our approach exclusively. We produced the best results with parametrisations based on band structures generated using the QSGW method, which is far more complete than the DFT. In principle, one could even use experimental band structures. In other words, by selecting the right reference we can develop a fast model of the electronic structure that can recreate features incorrectly represented by some *ab initio* methods. A good example of such a feature is the band-gap often underestimated by the DFT ([53]).

This brings us to the key characteristic of our method, namely how we can handle a diverse reference data-set without using inefficient multi-objective optimisation. In our approach, we explore the manifold of parameters associated with the electronic structure by addressing one set of targeted features (e.g. band structures) and generate a sufficiently large set of candidates. Later, this set is refined using another set of features (e.g. mechanical properties) to narrow down the search. Note that while we limited ourselves to two stages, in principle we could include also other properties (e.g. formation energies) and enclose the whole procedure in a loop. We believe that our method can be easier or faster when we compare it to approaches that involve building one performance measure that mixes multiple features. Naturally, this advantage refers to situations in which we would like to target many distinct features of a system.

Referring yet again to the fact that our models turned out be comparable, but not superior to other TB models, one can drive a conclusion that a complete *spd* basis does not introduce a significant improvement in terms of transferability.

Despite the issues encountered we still think that our approach is an interesting path to explore and multiple improvements can be introduced. For example, currently, we are investigating the possibility of improving the optimisation procedure by using all data generated so far. We aim to find an appropriate basis to describe the essential features of the band structures so we can find an inverse of the loss function. This would allow us to investigate which parameters can be changed without sacrificing the fit to the bands and test parameters from this subset to find ones that satisfy mechanical properties. As a result, it should be also possible to provide some approximation of the uncertainty of the band structure estimation.

Further improvement could involve testing a bigger variety of functions describing decays and the pair-potential, including cut-off radii in the optimisation stage or select functions that involve screening effects. There is also the somewhat controversial option of using machine-learning methods to find an exact interpolation of the optimal pair-potential, which might, but need not be, environmentally dependent. On top of the optimisation methodology, we also consider improvements in the reference data-set that could include calculation of some properties associated only with the electronic structure, such as magnetic susceptibility etc.

Finally, while we were aiming to create an automated workflow, we found that achieving this goal will still require a significant amount of development. Models presented in this paper were fine-tuned in a semi-automatic way (see figure 3), where we relied on our intuition to decide which parameters to modify. The hit-and-run tactics used in this stage could be automated. However, we were not able to find a measure that would allow us to quantify acceptable deviations (e.g. slight shift or scaling) in the fit of the band structures in the hope of achieving better mechanical properties. Furthermore, we lacked a proper treatment of band features well above Fermi level, where we would have to employ 4*d* orbitals to recreate all features. We speculate that these issues could be resolved by including in the reference information from Mulliken analysis and replacing the inefficient hit-and-run tactic with some form of adversarial machine learning.

Acknowledgments

We would like to kindly acknowledge The Engineering and Physical Sciences Research Council (EPSRC) for funding this Project (Ref. EP/P003591/1). We would like to also thank Dr. Dimitar Pashov and Dr. Mark Wenman for their valuable comments and continuous support. C P Race was funded by a University Research Fellowship of the Royal Society. Calculations

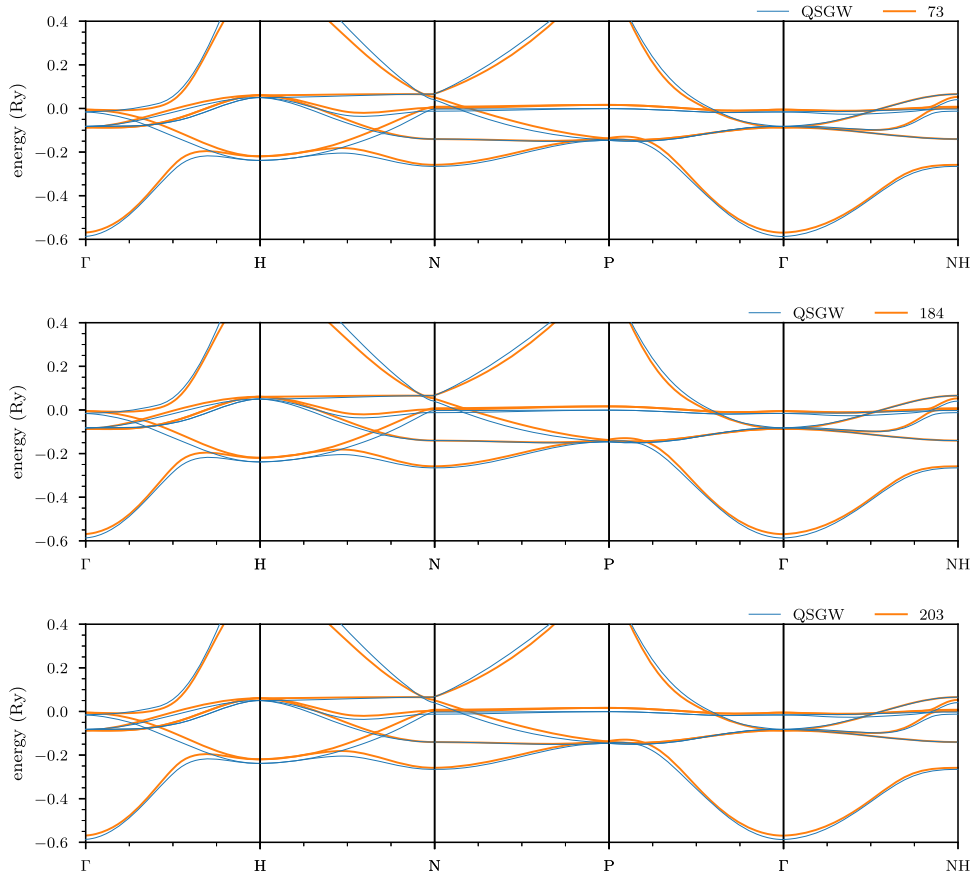

Figure A1. Band structures for bcc structure of Fe in the ferromagnetic state (majority spin). The blue line represents band structures plotted from electron density updated in the QSGW self-consistency cycle. The number indicates index of the TB model.

were performed on a computational cluster, maintained by the Computational Shared Facility, The University of Manchester.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Appendix A. Band plots

In this section we replot figure 6 to provide a better view for assessment of the quality of the fit (figures A1–A3).

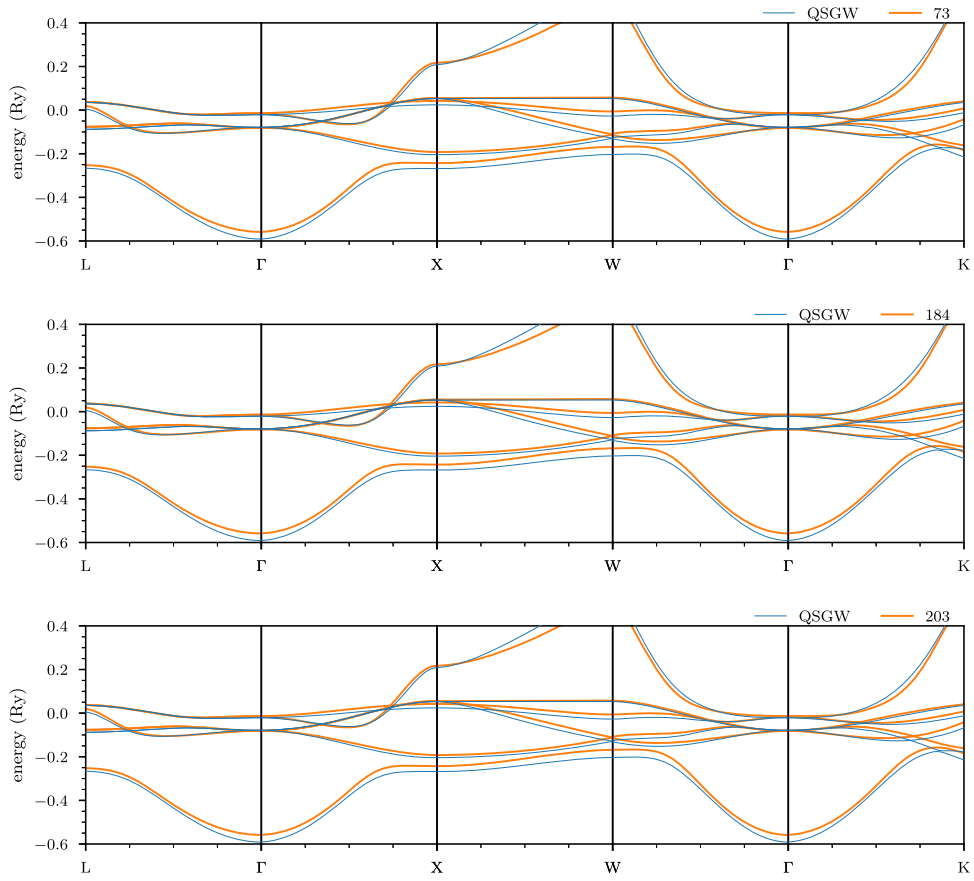

Figure A2. Band structures fcc structure of Fe in the ferromagnetic state (majority spin). Here, the lattice is under expansion 2% and .35% for models 73 and 184 respectively. The blue line represents band structures plotted from electron density updated in the QSGW self-consistency cycle. The number indicates index of the TB model.

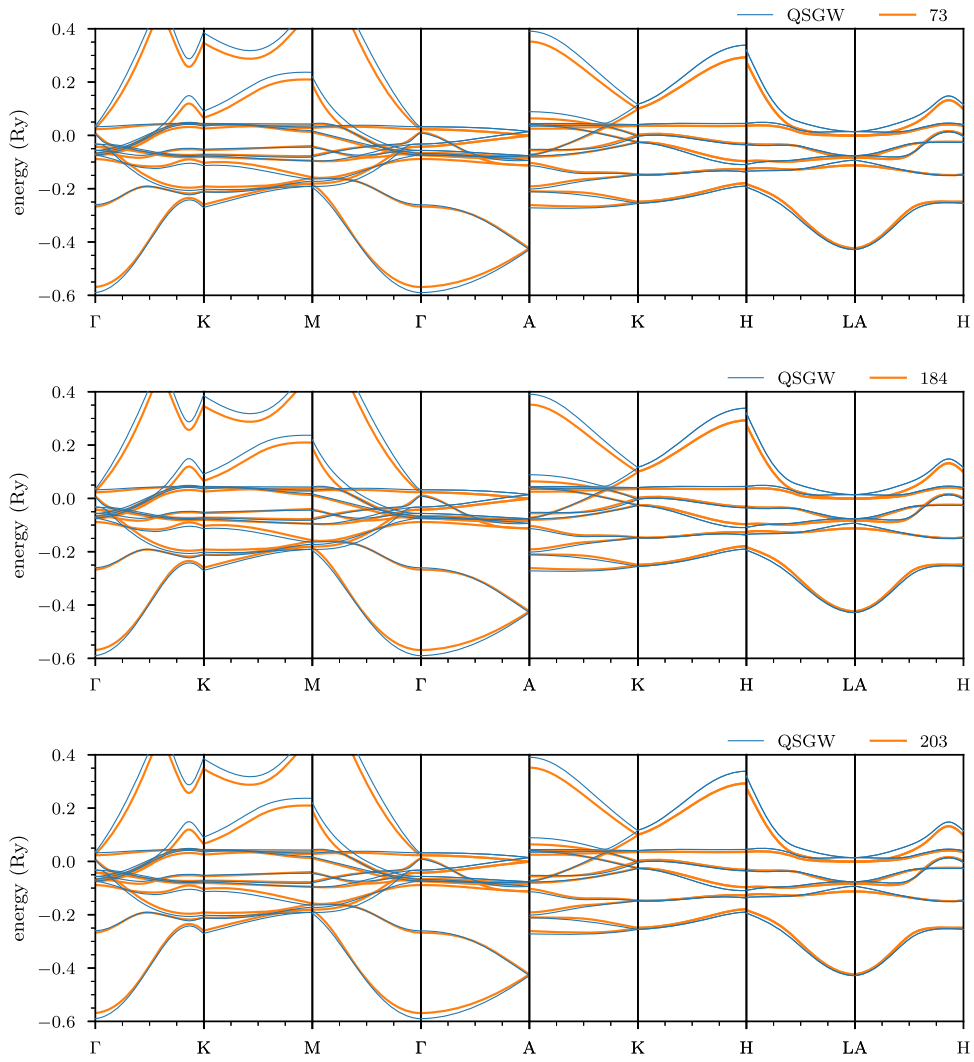

Figure A3. Band structures hcp structure of Fe in the ferromagnetic state (majority spin). The blue line represents band structures plotted from electron density updated in the QSGW self-consistency cycle. The number indicates index of the TB model.

ORCID iDs

Bartosz Barzdajn https://orcid.org/0000-0002-3081-4131
Christopher P Race https://orcid.org/0000-0002-9775-687X

References

[1] Slater J C and Koster G F 1954 Simplified LCAO method for the periodic potential problem *Phys. Rev.* **94** 1498–524

[2] Andersen O K and Jepsen O 1984 Explicit, first-principles tight-binding theory *Phys. Rev. Lett.* **53** 2571–4
[3] Cohen R E, Mehl M J and Papaconstantopoulos D A 1994 Tight-binding total-energy method for transition and noble metals *Phys. Rev.* B **50** 14694–7
[4] Mehl M J and Papaconstantopoulos D A 1996 Applications of a tight-binding total-energy method for transition and noble metals: elastic constants, vacancies, and surfaces of monatomic metals *Phys. Rev.* B **54** 4519–30
[5] Kohn W and Sham L J 1965 Self-consistent equations including exchange and correlation effects *Phys. Rev.* **140** A1133
[6] Paxton A T and Finnis M W 2008 Magnetic tight binding and the iron-chromium enthalpy anomaly *Phys. Rev.* B **77** 024428
[7] Stoner C 1938 Collective electron ferromagnetism *Proc. R. Soc.* A **165** 372–414
[8] Bacalis N C, Papaconstantopoulos D A, Mehl M J and Lach-Hab M 2001 Transferable tight-binding parameters for ferromagnetic and paramagnetic iron *Physica* B **296** 125–8
[9] Liu G, Nguyen-Manh D, Liu B G and Pettifor D G 2005 Magnetic properties of point defects in iron within the tight-binding-bond Stoner model *Phys. Rev.* B **71** 174115
[10] Paxton A T and Elsässer C 2010 Electronic structure and total energy of interstitial hydrogen in iron: tight-binding models *Phys. Rev.* B **82** 235125
[11] Madsen G K H, McEniry E J and Drautz R 2011 Optimized orthogonal tight-binding basis: application to iron *Phys. Rev.* B **83** 184119
[12] Hatcher N, Madsen G K H and Drautz R 2012 DFT-based tight-binding modeling of iron-carbon *Phys. Rev.* B **86** 155115
[13] Horsfield A and Kenny S D 1998 Efficient *ab initio* tight binding *Materials Research Society Symp. —Proc.* vol 491 pp 57–63
[14] Kotani T, Van Schilfgaarde M and Faleev S V 2007 Quasiparticle self-consistent GW method: a basis for the independent-particle approximation *Phys. Rev.* B **76** 165106
[15] See QSGW tutorial for magnetic bcc Fe: https://questaal.org/tutorial/gw/qsgw_fe/
[16] Paxton A T 2009 An introduction to the tight binding approximation—implementation by diagonalisation *Multiscale Simulation Methods in Molecular Sciences* vol 42 ed D Marx, J Grotendorst, N Attig and S Blügel (Jülich: Forschungszentrum Jülich) pp 145–76
[17] Sutton A P and Balluffi R W 1997 *Interfaces in Crystalline Materials* (Oxford: Clarendon)
[18] Finnis M 2010 *Interatomic Forces in Condensed Matter* (Oxford: Oxford University Press)
[19] Finnis M W, Paxton A T, Methfesselt M and van Schilfgaarde M 1997 Self-consistent tight-binding approximation including polarisable ions *Tight-Binding Approach to Computational Materials Science* pp 265–74
[20] Sutton A P, Finnis M W, Pettifor D G and Ohta Y 1988 The tight-binding bond model *J. Phys. C: Solid State Phys.* **21** 35–66
[21] Pashov D, Acharya S, Lambrecht W R L, Jackson J, Belashchenko K D, Chantis A, Jamet F and van Schilfgaarde M 2019 Questaal: a package of electronic structure methods based on the linear muffin-tin orbital technique *Comput. Phys. Commun.* **249** 107065
[22] Finnis M W, Paxton A T, Methfessel M and van Schilfgaarde M 1998 Crystal structures of zirconia from first principles and self-consistent tight binding *Phys. Rev. Lett.* **81** 5149–52
[23] Hubbard J 1963 Electron correlations in narrow energy bands *Proc. R. Soc.* A **276** 238–57
[24] Harrison W A 1985 Coulomb interactions in semiconductors and insulators *Phys. Rev.* B **31** 2121–32
[25] Foulkes W M C 2016 Tight-binding models and Coulomb interactions for s, p, and d electrons *Quantum Materials: Experiments and Theory* vol 6 ed E Pavarini, E Koch, J Van Den Brink and G Sawatzky (Jülich: Forschungszentrum Jülich) p 420
[26] Vekilov Y K, Krasilnikov O M, Lugovskoy A V and Lozovik Y E 2016 Higher-order elastic constants and megabar pressure effects of bcc tungsten: ab initio calculations *Phys. Rev.* B **94** 104114
[27] The atomic spheres approximation https://questaal.org/docs/code/asaoverview/ (accessed December 2019)
[28] Hansen N and Ostermeier A 1996 Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation *Proc. IEEE Int. Conf. on Evolutionary Computation* pp 312–7
[29] Hansen N, Akimoto Y and Baudis P 2019 CMA-ES/pycma on Github Zenodo https://doi.org/10.5281/zenodo.2559634

[30] Dufresne A, Ribeiro F and Tréglia G 2015 How to derive tight-binding spd potentials? Application to zirconium *J. Phys.: Condens. Matter.* **27** 336301
[31] Endres S C, Sandrock C and Focke W W 2018 A simplicial homology algorithm for Lipschitz optimisation *J. Global Optim.* **72** 181–217
[32] Virtanen P *et al* (SciPy 1.0 Contributors) 2019 SciPy 1.0—fundamental algorithms for scientific computing in Python (arXiv:1907.10121)
[33] Gao F and Han L 2012 Implementing the Nelder–Mead simplex algorithm with adaptive parameters *Comput. Optim. Appl.* **51** 259–77
[34] Nelder J A and Mead R 1965 A simplex method for function minimization *Comput. J.* **7** 308–13
[35] Goodwin L, Skinner A J and Pettifor D G 1989 Generating transferable tight-binding parameters: application to silicon *Europhys. Lett.* **9** 701–6
[36] Perdew J P, Burke K and Wang Y 1996 Generalized gradient approximation for the exchange–correlation hole of a many-electron system *Phys. Rev.* B **54** 16533–9
[37] Brugger K 1964 Thermodynamic definition of higher order elastic coefficients *Phys. Rev.* **133** 6
[38] Adams J J, Agosta D S, Leisure R G and Ledbetter H 2006 Elastic constants of monocrystal iron from 3 to 500 K *J. Appl. Phys.* **100** 113530
[39] Rayne J A and Chandrasekhar B S 1961 Elastic constants of iron from 4.2 to 300 K *Phys. Rev.* **122** 1714–6
[40] Hughes D S and Kelly J L 1953 Second-order elastic deformation of solids *Phys. Rev.* **92** 1145–9
[41] Choy M M, Hellwege K H and Hellwege A M 1979 *Elastic, Piezoelectric, Pyroelectric, Piezooptic, Electrooptic Constants, and Nonlinear Dielectric Susceptibilities of Crystals: Revised and Expanded Edition of Volumes III/1 and III/2* (*Numerical Data and Functional Relationships in Science and Technology, New Series: Crystal and Solid State Physics*) (Berlin: Springer)
[42] Pham Hieu H. and Çağın Tahir 2010 Lattice dynamics and second and third order elastic constants of iron at elevated pressures *Comput. Mater. Contin.* **16** 175–94
[43] Blaschke D N 2017 Averaging of elastic constants for polycrystals *J. Appl. Phys.* **122** 145110
[44] Alchagirov A B, Perdew J P, Boettger J C, Albers R C and Fiolhais C 2001 Energy and pressure versus volume: equations of state motivated by the stabilized jellium model *Phys. Rev.* B **63** 224115
[45] Kresse G and Furthmüller J 1996 Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set *Phys. Rev.* B **54** 11169–86
[46] Kresse G and Joubert D 1999 From ultrasoft pseudopotentials to the projector augmented-wave method *Phys. Rev.* B **59** 1758–75
[47] Wróbel J S, Nguyen-Manh D, Lavrentiev M Y, Muzyk M and Dudarev S L 2015 Phase stability of ternary fcc and bcc Fe–Cr–Ni alloys *Phys. Rev.* B **91** 024108
[48] Paxton A T and Elsässer C 2013 Analysis of a carbon dimer bound to a vacancy in iron using density functional theory and a tight binding model *Phys. Rev.* B **87** 224110
[49] Dragoni D, Daff T D, Csányi G and Marzari N 2018 Achieving DFT accuracy with a machine-learning interatomic potential: thermomechanics and defects in bcc ferromagnetic iron *Phys. Rev. Mater.* **2** 13808
[50] Bartók A P, Payne M C, Kondor R and Csányi G 2010 Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons *Phys. Rev. Lett.* **104** 136403
[51] Messina L, Chang Z and Olsson P 2013 *Ab initio* modelling of vacancy-solute dragging in dilute irradiated iron-based alloys *Nucl. Instrum. Methods Phys. Res.* B **303** 28–32
[52] Whiting T M, Burr P A, King D J M and Wenman M R 2019 Understanding the importance of the energetics of Mn, Ni, Cu, Si and vacancy triplet clusters in bcc Fe *J. Appl. Phys.* **126** 115901
[53] van Schilfgaarde M, Kotani T and Faleev S 2006 Quasiparticle self-consistent GW theory *Phys. Rev. Lett.* **96** 226402